

Thresholding methods to estimate the copula density

F. Autin, E. Le Pennec, K. Tribouley

Universités Aix-Marseille 1, Paris 7 et Paris 10

Abstract

This paper deals with the problem of the multivariate copula density estimation. Using wavelet methods we provide two shrinkage procedures based on thresholding rules for which the knowledge of the regularity of the copula density to be estimated is not necessary. These methods, said to be adaptive, are proved to perform very well when adopting the minimax and the maxiset approaches. Moreover we show that these procedures can be discriminated in the maxiset sense. We produce an estimation algorithm whose qualities are evaluated thanks some simulation. Last, we propose a real life application for financial data.

Key words: copula density, wavelet method, thresholding rules, minimax theory, maxiset theory.

AMS Subject Classification: 62G10, 62G20, 62G30.

1 Introduction

Recently a new tool has appeared (principally for risk management, for instance in finance, insurance, climatology, hydrology etc) to model the structure of dependance of the data. Let us start by recalling the seminal result of Sklar [17].

Theorem 1 (Sklar (1959)) *Let $d \geq 2$ and H be a d -variate distribution function. If, for $m = 1, \dots, d$, any margin F_m of H is continuous, there exists a unique d -variate function C with uniform margins $\mathcal{U}_{[0,1]}$ such that*

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

The repartition function C is called **the copula** associated to the distribution H . The interest of the Sklar Theorem is the following: it is possible to study

separately the laws of the coordinates $X^m, m = 1, \dots, d$, of any vector X whose the law is H and the dependance between the coordinates.

The **Copulas model** has been extensively studied in a parametrical framework for the distribution function C . Large classes of Copulas, such as the elliptic family, which contains the Gaussian Copulas and the Student Copula, and the archimedian family, which contains the Gumbel Copula, the Clayton Copula and the Frank Copulas, have been identified. Mainly, people have worked in two directions. Firstly, an important activity has concerned **the modeling** in view to find new Copulas and methodologies to simulate data coming from these new Copulas. Secondly, usual **statistical inference** (estimation of the parameters, goodness-of-fit test, etc) has been developed using the copulas. As usual, the nonparametric point of view is useful when no a priori model of the phenomenon is specified. For the practitioners, the nonparametrical estimators could be seen as a **benchmark** allowing to specify the model, comparing with the available parametrical families. This explains the success of the nonparametric estimator of the copula. Unfortunately, the practitioners have difficulties to analyze the graphes concerning the distribution functions (when $d = 2$). Generally, they try to make comments observing the scatter plot of $\{(X_i, Y_i), i = 1, \dots, n\}$ (or $\{(R_i, S_i), i = 1, \dots, n\}$ where R, S are the rank statistics of X, Y). Therefore a density estimator could be very advantageous.

We propose procedures to estimate the density associated to the copula C . This density denoted c is called the **copula density**. The copula density model that we consider is the following. Let us give a n -sample $(X_1^1, \dots, X_1^d), \dots, (X_n^1, \dots, X_n^d)$ of independent data admitting the same distribution H (and the same density h) as (X^1, \dots, X^d) . We denote F_1, \dots, F_d the margins of the coordinates of the vector (X^1, \dots, X^d) . We are interested in the estimation of the density copula c which is defined as the derivative (if it exists) of the copula distribution

$$c(u_1, \dots, u_d) = \frac{h(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \dots f_d(F_d^{-1}(u_d))}$$

where $F_p^{-1}(u_p) = \inf\{x \in R : F_p(x) \geq u_p\}$, $1 \leq p \leq d$ and $u = (u_1, \dots, u_d) \in [0, 1]^d$. This model is a very classical density model but the direct observations $(U_i^1 = F_1(X_i^1), \dots, U_i^d = F_d(X_i^d))$ for $i = 1, \dots, n$ of the copula C are not available, because the margins are generally unknown. Observe that a similar problem could be the nonparametric regression model with random design. This model has been studied in Kerkycharian and Picard [15] using **warped wavelet families**.

In this paper, we focus on wavelet methods. We emphasize that these methods are very well appropriated because the localization properties of the wavelet allow us to give a sharp analysis of the behavior of the copula density near the

border of $[0, 1]^d$. A great advantage of the wavelet methods in statistics is to provide adaptive procedures in the sense that they automatically adapt to the regularity of the object to be estimated. The problem is a little bit different in the case of the copula. So far, the practitioners use copula densities which are regular. Nevertheless, one of the practical problem is due to the fact that the copula densities are generally concentrated towards the borders of $[0, 1]^d$ (see Part 6 where simulations are presented). It is very difficult to provide a good estimation when the considered density admits significant jumps (which is the case of the copula density when its support is extended). Observe that the lack of data behind the jumps produce poor estimators and this is just where we are interested by the estimation. The thresholding methods use multifrequency levels depending of the place where the estimator is computed. Consequently, they are really fitted for the problem to estimate the copula density and we think that they outperform the classical linear methods like kernel estimators.

We present in this paper the **Local Thresholding Method** and the **Global (or Block) Thresholding Method**. These methods has been first studied by Donoho et al. ([6], [7]) and Kerkycharian et al. [13]. We prove that the theoretical properties of these procedures with respect to the quadratic risk are as good in the copula density model as in the usual density model (when direct data are available). The good behavior of the usual procedures in the copula density model is also observed when linear procedures are considered (see Genest et al. [9]) and for test problems (see Gayraud and Tribouley [8]). To give an entire overview on the minimax properties of the copula density estimators, we explore the maxiset properties of both procedures (and also of the linear procedure). Again, we prove that the fact that no direct observations are available does not affect the properties of the wavelet procedures: as in the standard density model, the local thresholding method of estimation outperforms the others one.

Next, we provide an important practical study. Another advantage of the wavelet procedures is their remarkable facility of use. Indeed, a large number of implementation of the fast wavelet transform exist and it is just a matter of choice to select one in a given programming language. They can be used almost directly after a simple warping of the original samples. Nevertheless, as noticed previously in the case of the copula density, as most of the pertinent information is located near the border of $[0, 1]^d$, the handling of the boundaries should be done with a special care. It is one of the result of the article to show that an inappropriate handling, such as the one proposed by default by a lot of implementation, could cause the wavelet method to fail. The symmetrization/periodization process proposed in this paper is described. To reduce the gridding effect of the wavelet, we further propose to incorporate some limited translation invariance in the method. We first comment our results for simulated data in the case of the usual parametrical copula families and then we present an application on financial data: we propose a method

to estimate the copula density imposing that the "true" copula belongs to a target parametrical family and using the nonparametrical estimator as a benchmark.

The paper is organized as follows. Section 2 deals with the wavelet setting so as to describe the multidimensional wavelet basis used in the sequel. Section 3 aims at describing both thresholding procedures of estimation for which performances shall be studied in Section 4 with the minimax approach and in Section 5 with the maxiset approach. Section 6 is devoted to the practical results. Proofs of main theorems are given in Section 7 and shall need proposition and technical lemmas proved in Appendix.

2 Wavelet setting

Let ϕ and ψ be respectively a scaling function and an associated wavelet function. We assume that these functions are compactly supported on $[0, L]$ for some $L \geq 1$. See for example the Daubechies's wavelets (see Daubechies [5]). For any univariate function $h(\cdot)$, we denote by $h_{j,k}(\cdot)$ the function $2^{j/2}h(2^j \cdot -k)$ where $j \in \mathbb{N}$ and $k \in \mathbb{Z}$. In the sequel, we use wavelet expansions for multivariate functions. We build a multivariate wavelet basis as follows:

$$\begin{aligned}\phi_{j,k}(x_1, \dots, x_d) &= \phi_{j,k_1}(x_1) \dots \phi_{j,k_d}(x_d), \\ \psi_{j,k}^\epsilon(x_1, \dots, x_d) &= \prod_{m=1}^d \phi_{j,k_m}^{1-\epsilon_m}(x_m) \psi_{j,k_m}^{\epsilon_m}(x_m),\end{aligned}$$

for all $\epsilon = (\epsilon_1, \dots, \epsilon_d) \in S_d = \{0, 1\}^d \setminus \{(0, \dots, 0)\}$. We keep the same notation for univariate scaling function and multivariate scaling function; the subscripts $k = (k_1, \dots, k_d)$ indicate the number of components. For any $j_0 \in \mathbb{N}$, the set $\{\phi_{j_0 k}, \psi_{j,k}^\epsilon \mid j \geq j_0, k \in \mathbb{Z}^d, \epsilon \in S_d\}$ is an orthonormal basis of $L_2(\mathbb{R}^d)$ and the expansion of any real function h of $L_2(\mathbb{R}^d)$ is given by:

$$\forall x \in \mathbb{R}^d, h(x) = \sum_{k \in \mathbb{Z}^d} h_{j_0, k} \phi_{j_0, k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}^d} \sum_{\epsilon \in S_d} h_{j, k}^\epsilon \psi_{j, k}^\epsilon(x),$$

where, for any $(j, k, \epsilon) \in \mathbb{N} \times \mathbb{Z}^d \times S_d$, the wavelet coefficients are

$$h_{j_0, k} = \int_{\mathbb{R}^d} h(x) \phi_{j_0, k}(x) dx \text{ and } h_{j, k}^\epsilon = \int_{\mathbb{R}^d} h(x) \psi_{j, k}^\epsilon(x) dx.$$

Roughly speaking, the expansion of the analyzed function on the wavelet basis splits into the "trend" at the level j_0 and the sum of the "details" for all the larger levels j , $j \geq j_0$.

3 Estimation procedures

Assuming that the copula density c belongs to L_2 , we present wavelet procedures of estimation. Motivated by the wavelet expansion, we first estimate the coefficients of the copula density on the wavelet basis. Observe that, for any d -variate function Φ

$$E_c(\Phi(U_1, \dots, U_d)) = E_h(\Phi(F_1(X^1), \dots, F_d(X^d)))$$

which means that the wavelet coefficients $c_{j_0,k}, c_{j,k}^\epsilon$ of the copula density c on the wavelet basis $\{\phi_{j_0,k}, \psi_{j,k}^\epsilon \mid j \geq j_0, k \in \mathbb{Z}^d, \epsilon \in S_d\}$ are equal to the coefficients of the joint density h on the warped wavelet family

$$\{\phi_{j_0,k}(F_1(\cdot), \dots, F_d(\cdot)), \psi_{j,k}^\epsilon(F_1(\cdot), \dots, F_d(\cdot)) \mid j \geq j_0, k \in \mathbb{Z}^d, \epsilon \in S_d\}.$$

As usual, standard empirical coefficients are

$$\widehat{c_{j_0,k}} = \frac{1}{n} \sum_{i=1}^n \phi_{j_0,k}(F_1(X_i^1), \dots, F_d(X_i^d))$$

and

$$\widehat{c_{j,k}^\epsilon} = \frac{1}{n} \sum_{i=1}^n \psi_{j,k}^\epsilon(F_1(X_i^1), \dots, F_d(X_i^d)). \quad (1)$$

Since no direct observation $(F_1(X_i^1), \dots, F_d(X_i^d))$ is usually available, we propose to replace this one with the pseudo observation $(\widehat{F}_1(X_i^1), \dots, \widehat{F}_d(X_i^d))$ where $\widehat{F}_1, \dots, \widehat{F}_d$ are the empirical distribution functions associated with the margins. The empirical coefficients are

$$\widetilde{c_{j_0,k}} = \frac{1}{n} \sum_{i=1}^n \phi_{j_0,k}(\widehat{F}_1(X_i^1), \dots, \widehat{F}_d(X_i^d)) = \frac{1}{n} \sum_{i=1}^n \phi_{j_0,k}\left(\frac{R_i^1}{n}, \dots, \frac{R_i^d}{n}\right)$$

and

$$\widetilde{c_{j,k}^\epsilon} = \frac{1}{n} \sum_{i=1}^n \psi_{j,k}^\epsilon(\widehat{F}_1(X_i^1), \dots, \widehat{F}_d(X_i^d)) = \frac{1}{n} \sum_{i=1}^n \psi_{j,k}^\epsilon\left(\frac{R_i^1}{n}, \dots, \frac{R_i^d}{n}\right)$$

where R_i^p denotes the rank of X_i^p for $p = 1, \dots, d$

$$R_i^p = \sum_{l=1}^n \mathbf{1}\{X_l^p < X_i^p\}.$$

According to the fact that the wavelet basis is compactly supported, the sum over the indices k, ϵ is finite and is taken over $2^d L^d 2^{dj}$ terms. In the sequel, to simplify the notations, we omit the bounds of variation of the indices k, ϵ .

The most natural way to estimate the density c is to reconstruct the density thanks to the estimated coefficients. For any indices (j_n, J_n) such that $j_n \leq J_n$, we consider the very general family of **truncated estimators** of c defined by

$$\widetilde{c}_T := \widetilde{c}_T(j_n, J_n) = \sum_k \widetilde{c}_{j_n, k} \phi_{j_n, k} + \sum_{j=j_n}^{J_n} \sum_{k, \epsilon} \omega_{j, k}^\epsilon \widetilde{c}_{j, k}^\epsilon \psi_{j, k}^\epsilon, \quad (2)$$

where $\omega_{j, k}^\epsilon \in \{0, 1\}$ for any (j, k, ϵ) . It is intuitive that the more regular is the function c , the smallest are the details and then that a good approximation for the function c is the trend at a level j_n large enough. Such **linear procedures**

$$\widetilde{c}_L := \widetilde{c}_L(j_n) = \sum_k \widetilde{c}_{j_n, k} \phi_{j_n, k} \quad (3)$$

consisting to put $\omega_{j, k}^\epsilon = 0$ for any (j, k, ϵ) have been considered in Genest et al. [9] where a discussion on the choice of j_n is done. It is also possible to consider other truncated procedures as **non linear procedures**: the local thresholding procedure consists to "kill" individually the small estimated coefficients considering that they do not give any information. Let $\lambda_n > 0$ be a thresholding level and (j_n, J_n) be the level indices. For $\epsilon \in S_d$ and j varying between j_n and J_n , the hard local threshold estimators of the wavelet coefficients are

$$\widetilde{d}_{j, k}^\epsilon = \widetilde{c}_{j, k}^\epsilon \mathbf{1}\{|\widetilde{c}_{j, k}^\epsilon| > \lambda_n\},$$

leading to the hard local threshold procedure

$$\widetilde{c}_{HL} := \widetilde{c}_{HL}(j_n, J_n) = \sum_k \widetilde{c}_{j_n, k} \phi_{j_n, k} + \sum_{j=j_n}^{J_n} \sum_{k, \epsilon} \widetilde{d}_{j, k}^\epsilon \psi_{j, k}^\epsilon. \quad (4)$$

It is also possible to decide to "kill" all the coefficients of the level j if the information given by this level is too small. For a thresholding level $\lambda_n > 0$ and for levels j varying between j_n and J_n , we define the hard global threshold estimates of the wavelet coefficients as follows

$$\widetilde{e}_{j, k}^\epsilon = \widetilde{c}_{j, k}^\epsilon \mathbf{1}\left\{\sum_k |\widetilde{c}_{j, k}^\epsilon|^2 > L^d 2^{dj} \lambda_n^2\right\},$$

leading to the hard global threshold procedure

$$\widetilde{c}_{HG} := \widetilde{c}_{HG}(j_n, J_n) = \sum_k \widetilde{c}_{j_n, k} \phi_{j_n, k} + \sum_{j=j_n}^{J_n} \sum_{k, \epsilon} \widetilde{e}_{j, k}^\epsilon \psi_{j, k}^\epsilon. \quad (5)$$

The non linear procedures given in (4) and (5) depend on the levels indices (j_n, J_n) and on the thresholding level λ_n to be chosen by the user. In the next part, we explain how to determine these parameters such a way that the associated procedures achieve optimality properties. We need then to define a criterion to measure the performance of our procedure.

4 Minimax Results

4.1 Minimax approach

A well-known way to analyze the performances of procedures of estimation is the **minimax theory** which has been extensively developed since the 1980-ies. In the minimax setting, the practitioner chooses a loss function $\ell(\cdot)$ to measure the loss due to the studied procedure and a functional space \mathcal{F} where the unknown object to estimated is supposed to belong. The choice of this space is important because the first step (when the minimax theory is applied) consists to compute the **minimax risk** associated to this functional space. This quantity

$$R_n(\mathcal{F}) = \inf_{\widetilde{c}} \sup_{c \in \mathcal{F}} E \ell(\widetilde{c} - c)$$

(where the infimum is taken over all the estimators of c) is a lower bound giving the best rate achievable on the space \mathcal{F} . When a statistician proposes an estimation procedure for functions belonging to \mathcal{F} , he has to evaluate the risk of his procedure and compare this upper bound with the minimax risk. If the rates coincide, the procedure is **minimax optimal** on the space \mathcal{F} . A lot of minimax results for many statistical models and many families of functional spaces as Sobolev spaces, Holder spaces, and others as the family of Besov spaces have been now established (see for instance Ibragimov and Khasminski [11] or Kerkyacharian and Picard [12]).

4.2 Besov spaces

Since we deal with wavelet methods, it is very natural to consider Besov spaces as functional spaces because they are characterized in term of wavelet coefficients as follows

Definition 1 (Strong Besov spaces) *For any $s > 0$, a function c belongs*

to the Besov space $\mathcal{B}_{2\infty}^s$ if and only if its sequence of wavelet coefficients $c_{j,k}^\epsilon$ satisfies

$$\sup_{J \geq 0} 2^{2Js} \sum_{j > J} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 < \infty.$$

An advantage of the Besov spaces $\mathcal{B}_{2\infty}^s$ is to provide a useful tool to classify wavelet decomposed signals according to their regularity and sparsity properties (see for instance Donoho and Johnstone [6]). Last, it is well known that the minimax risk measured with the quadratic loss on this space is

$$\sup_n \inf_{\tilde{c}} \sup_{c \in \mathcal{B}_{2\infty}^s} n^{\frac{2s}{2s+d}} E \|\tilde{c} - c\|_2^2 < \infty$$

where the infimum is taken over any estimator of the density c .

4.3 Optimality

Let us focus on the quadratic loss function. Choosing a wavelet regular enough and when $d = 2$, Genest et al. [9] prove that the linear procedure $\tilde{c}_L = \tilde{c}_L(j_n)$ defined in (3) provides an optimal estimator on the Besov space $\mathcal{B}_{2\infty}^s$ for some fixed $s > 0$ as soon as j_n is chosen as follows:

$$2^{j_n-1} < n^{\frac{1}{2s+d}} \leq 2^{j_n}.$$

On a practical point of view, this result is not completely satisfying because the optimal procedure depends on the regularity s of the density which is generally unknown. To avoid this drawback, many works in the nonparametric setting as in Cohen et al. [12], in Kerkycharian and Picard [14] inspired from Donoho and Johnstone's studies on shrinkage procedures (see for instance [6]) build **adaptive procedures** which means that they do not depend on some a priori information about the unknown density. For instance, note that the thresholding procedures described in (4) and (5) are clearly adaptive. The following theorem (which proof is a direct consequence of Theorem 4 established in the next section by considering some inclusion spaces properties) gives results on their rates

Theorem 2 *Let us consider a continuously differentiable wavelet function and let $s > 0$. Let us choose the integers j_n and J_n and the real λ_n such that*

$$2^{j_n-1} < (\log(n))^{1/d} \leq 2^{j_n}, \quad 2^{J_n-1} < \left(\frac{n}{\log n} \right)^{1/d} \leq 2^{J_n}, \quad \lambda_n = \sqrt{\frac{\kappa \log(n)}{n}}$$

for some $\kappa > 0$. Let \tilde{c} be either the hard local thresholding procedure $\tilde{c}_{HL}(j_n, J_n)$ or the hard global thresholding procedure $\tilde{c}_{HG}(j_n, J_n)$. Then, as soon as κ is

large enough, we get

$$c \in \mathcal{B}_{2\infty}^s \cap L_\infty \implies \sup_n \left(\frac{n}{\log(n)} \right)^{\frac{2s}{2s+d}} E \|\tilde{c} - c\|_2^2 < \infty.$$

We immediately deduce

Corollary 4.1 *The hard local thresholding procedure $\widetilde{c_{HL}}$ and the hard global thresholding procedure $\widetilde{c_{HG}}$ are adaptive near optimal up to a logarithmic term (that is the price to pay for adaptation) on the Besov spaces $\mathcal{B}_{2\infty}^s$ for the quadratic loss function, according to the minimax point of view.*

Observe that if $s > d/2$ then the assumption $c \in \mathcal{B}_{2\infty}^s \cap L_\infty$ in Theorem 2 could be replaced by $c \in \mathcal{B}_{2\infty}^s$ because in this particular case $\mathcal{B}_{2\infty}^s \subsetneq L_\infty$.

4.4 Criticism on the minimax point of view

On a practical point of view, the first drawback of the minimax theory is the necessity to fix a functional space. Corollary 4.1 establishes that no procedure could be better on the space $\mathcal{B}_{2\infty}^s$ than our hard thresholding procedures (up to the logarithm factor) but this space is generally an abstraction for the practitioner. Moreover, this one knows that his optimal procedure achieves the best rate but he does not know this rate. An answer to this drawback is given by Lepski by introducing the concept of the random normalized factor (see Hoffmann and Lepski [10]).

Secondly, he has the choice between both procedures since Theorem 2 establishes that the hard thresholding procedures (local and global) have the same performances when dealing with the minimax point of view. Nevertheless, a natural question arises here: is it possible to compare thresholding procedures for estimation of the copula density between themselves? To answer to these remarks, we propose to explore in the next part the **maxiset approach**.

5 Maxiset Results

5.1 Maxiset approach

The maxiset point of view has been developed by Cohen et al. [4] and is inspired from recent studies in the approximation theory field. This approach

aims at providing a new way to analyze the performances of the estimation procedures. Contrary to the minimax setting, the maxiset approach consists in finding the maximal space of functions (called the maxiset) for which a given procedure attains a prescribed rate of convergence. According to this, the maxiset setting is not so far from the minimax one. Nevertheless it seems to be more optimistic than the minimax point of view in the sense that the maxiset approach points out **all** the functions **well-estimated** by a fixed procedure at a given accuracy.

In the sequel, we say that a functional space, namely $\mathcal{MS}(\tilde{c}, r_n)$, is the maxiset of a fixed estimation procedure \tilde{c} associated with the rate of convergence r_n if and only if the following equivalence is satisfied

$$\sup_n r_n^{-1} E \|\tilde{c} - c\|_2^2 < \infty \iff c \in \mathcal{MS}(\tilde{c}, r_n).$$

Notice that the considered loss function is again the quadratic one. As a first consequence of adopting the maxiset point of view, we observe that if \tilde{c} is an estimator of c achieving the minimax rate of convergence v_n on a functional space \mathcal{V} , then \mathcal{V} is included in the maxiset of \tilde{c} associated with the rate v_n but is not necessarily the same. Therefore, it could be possible to distinguish between both optimal minimax procedures: for the same target rate, the best procedure is the procedure admitting the largest maxiset.

Recently many papers based on the maxiset approach have arisen when considering the white noise model. For instance it has been proved in Autin et al. [2] that the hard local thresholding procedure is the best one, in the maxiset sense, among a large family of shrinkage procedures, called the **elitist rules**, composed of all the wavelet procedures only using in their construction empirical wavelet coefficients with absolute value larger than a prescribed quantity. This optimality has been already pointed out by Autin [1] in density estimation who has proved that **weak Besov spaces** are the saturation spaces of elitist procedures.

5.2 Weak Besov spaces

To model the sparsity property of functions, a very convenient and natural tool consists in introducing the following particular class of Lorentz spaces, namely **weak Besov spaces**, that are in addition directly connected to the estimation procedures considered in this paper. We give definitions of the weak Besov spaces depending on the wavelet basis. However, as is established in Meyer [16] and Cohen et al. [4], most of them have also different definitions proving that this dependence in the basis is not crucial at all.

Definition 2 (Local weak Besov spaces) For any $0 < r < 2$, a function $c \in L_2([0, 1]^d)$ belongs to the local weak Besov space $\mathcal{W}_L(r)$ if and only if its sequence of wavelet coefficients $c_{j,k}^\epsilon$ satisfies the following equivalent properties:

- $\sup_{0 < \lambda \leq 1} \lambda^{r-2} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \lambda\} < \infty$,
- $\sup_{0 < \lambda \leq 1} \lambda^r \sum_{j \geq 0} \sum_{k, \epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \lambda\} < \infty$.

Definition 3 (Global weak Besov spaces) For any $0 < r < 2$, a function $c \in L_2([0, 1]^d)$ belongs to the global weak Besov space $\mathcal{W}_G(r)$ if and only if its sequence of wavelet coefficients $c_{j,k}^\epsilon$ satisfies the following equivalent properties:

- $\sup_{0 < \lambda \leq 1} \lambda^{r-2} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{\sum_k (c_{j,k}^\epsilon)^2 \leq 2^{dj} \lambda^2\} < \infty$,
- $\sup_{0 < \lambda \leq 1} \lambda^r \sum_{j \geq 0} 2^{dj} \sum_{\epsilon} \mathbf{1}\{\sum_k (c_{j,k}^\epsilon)^2 > 2^{dj} \lambda^2\} < \infty$.

The equivalences between the properties used in the definition of global weak Besov spaces and in the definition of local weak Besov spaces can be proved using same technics as those proposed in Cohen et al. [4].

We prove in Section 7.3 the following link between the global weak Besov space and the local weak Besov space

Proposition 1 For any $0 < r < 2$, we get $\mathcal{W}_G(r) \subsetneq \mathcal{W}_L(r)$.

Last, we give an upper bound for the estimation expected error (when the thresholding procedures are used) in the standard density model (i.e. when direct observations $(U_i^1, \dots, U_i^d), i = 1, \dots, n$ are available). Notice that this result is stronger than the result given in the minimax part because the functional assumption is weaker.

Theorem 3 Let \bullet be either L or G and $\widehat{c_{H\bullet}}$ be the estimator $\widehat{c_{H\bullet}}$ built in the same way as $\widehat{c_{H\bullet}}$ but with the sequence of coefficients $\widehat{c_{j,k}^\epsilon}$ defined in (1). Let $s > 0$ and assume that

$$c \in \mathcal{W}_\bullet\left(\frac{2d}{2s+d}\right) \cap \mathcal{B}_{2,\infty}^{\frac{ds}{2s+d}}.$$

Then, there exists some $K > 0$ such that for any n

$$E\|\widehat{c_{H\bullet}} - c\|_2^2 \leq K \left(\frac{\log(n)}{n} \right)^{\frac{2s}{2s+d}}$$

There is no difficulty to prove this result by using the same technics as in Cohen et al. [4] where the proof is done for $d = 1$ for classical thresholding rules in general models. Nevertheless, for the interested reader, a detailed proof can be found in Autin et al. [3].

5.3 Performances and comparison of our procedures

In this section, we study the maxiset performances of the linear procedure and of the thresholding procedures described in Section 1. We focus on the optimal minimax procedures which means that we consider the following choices of parameters:

$$2^{j_n^*-1} < \left(\frac{n}{\log(n)} \right)^{\frac{1}{2s+d}} \leq 2^{j_n^*}, \quad \lambda_n = \sqrt{\frac{\kappa \log(n)}{n}} \text{ for some } \kappa > 0,$$

$$2^{j_n-1} < (\log(n))^{1/d} \leq 2^{j_n}, \quad 2^{J_n-1} < \left(\frac{n}{\log(n)} \right)^{1/d} \leq 2^{J_n}.$$

Let us fix $s > 0$. we choose to focus on the (near) minimax rate $r_n = (n^{-1} \log(n))^{2s/(2s+d)}$ achieved on the space $B_{2\infty}^s$. The following theorem exhibits the maxisets of the three procedures associated with the same target rate r_n .

Theorem 4 *Let $s > 0$, and assume that $c \in L_\infty$. For a large choice of κ , we have*

$$\sup_n \left(\frac{n}{\log(n)} \right)^{\frac{2s}{2s+d}} E\|\widetilde{c_L} - c\|_2^2 < \infty \iff c \in \mathcal{B}_{2\infty}^s, \quad (6)$$

$$\sup_n \left(\frac{n}{\log(n)} \right)^{\frac{2s}{2s+d}} E\|\widetilde{c_{HL}} - c\|_2^2 < \infty \iff c \in \mathcal{B}_{2\infty}^{\frac{ds}{2s+d}} \cap \mathcal{W}_L\left(\frac{2d}{2s+d}\right), \quad (7)$$

$$\sup_n \left(\frac{n}{\log(n)} \right)^{\frac{2s}{2s+d}} E\|\widetilde{c_{HG}} - c\|_2^2 < \infty \iff c \in \mathcal{B}_{2\infty}^{\frac{ds}{2s+d}} \cap \mathcal{W}_G\left(\frac{2d}{2s+d}\right). \quad (8)$$

It is important to notice that, according to Theorem 3, the fact that direct observations are not available does not affect the maxiset performances of our procedures. As remarked by Autin et al. [3] we have clearly

$$\mathcal{B}_{2\infty}^s \subsetneq \mathcal{B}_{2\infty}^{\frac{ds}{2s+d}} \cap \mathcal{W}_G(\frac{2d}{2s+d}).$$

We deduce by using Proposition 1 that both thresholding estimators considered in Theorem 2 achieve the minimax rate (up to the logarithmic term) on a larger functional space than $\mathcal{B}_{2\infty}^s$ which is the required space in the minimax approach. In particular, Theorem 2 is proved. We propose now to discriminate these procedures by comparing their maxisets. Thanks to Theorem 4 and applying the inclusion property given in Proposition 1, we prove in Section 7.3 the following corollary

Corollary 5.1 *Let $s > 0$ and let us consider the target rate*

$$r_n = \left(\frac{\log(n)}{n} \right)^{\frac{2s}{2s+d}}.$$

We get

$$\mathcal{MS}(\widetilde{c}_L, r_n) \subsetneq \mathcal{MS}(\widetilde{c}_{HG}, r_n) \subsetneq \mathcal{MS}(\widetilde{c}_{HL}, r_n).$$

Hence, in the maxiset point of view and when the quadratic loss is considered, the thresholding rules outperform the linear procedure. Moreover, the hard local thresholding estimator \widetilde{c}_{HL} appears to be the best estimator among the considered procedures since it strictly outperforms the hard global thresholding estimator \widetilde{c}_{HG} .

6 Applied results

This part is not only an illustration of the theoretical part. First, we explain the considered algorithm with several numerical possibilities to overcome any drawback. Next, we test the qualities of our methodology with some simulation and we define the best choices among our propositions. Last, we apply the chosen procedure for financial data.

6.1 Algorithms

The estimation algorithms are described here for $d = 2$ for sake of simplicity but their extension in any other dimension is straightforward. We therefore

assume that a sequence $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ of n samples is given.

All estimators proposed in this paper can be summarized in a 7 steps algorithm:

- (1) Rank the X_i, Y_i with

$$R_i = \sum_{l=1}^n \mathbf{1}_{X_l < X_i} \text{ and } S_i = \sum_{l=1}^n \mathbf{1}_{Y_l < Y_i}.$$

- (2) Compute the maximal scale index $J_n = \lfloor \frac{1}{2} \log_2(\frac{n}{\log n}) \rfloor$.
- (3) Compute the empirical scaling coefficients at the maximal scale index J_n :

$$\widetilde{c_{J_n, k_1, k_2}} = \frac{1}{n} \sum_{i=1}^n \phi_{J_n, k_1, k_2}(\frac{R_i}{n}, \frac{S_i}{n}) \quad \text{for } 1 \leq k_1 \leq 2^{J_n} \text{ and } 1 \leq k_2 \leq 2^{J_n}.$$

- (4) Compute the empirical wavelet coefficients $\widetilde{c_{j, k_1, k_2}^\epsilon}$ from this scaling coefficients with the fast 2D wavelet transform algorithm.
- (5) Threshold these coefficients according to the global thresholding rule or the local thresholding rule to obtained the estimated wavelet coefficients $\widetilde{e_{j, k_1, k_2}^\epsilon}$ or $\widetilde{d_{j, k_1, k_2}^\epsilon}$.
- (6) Compute the estimated scaling coefficients $\widetilde{c_{J_n, k_1, k_2}}$ at scale index J_n by the fast 2D wavelet inverse transform algorithm.
- (7) Construct the estimated copula density \tilde{c} by the formula

$$\tilde{c} = \sum_{k_1, k_2} \widetilde{c_{J_n, k_1, k_2}} \phi_{J_n, k_1, k_2} \quad .$$

Unfortunately only the steps (1), (2) and (5) are as straightforward as they seem. In all the other steps, one has to tackle with two issues: the handling of the boundaries and the fact that the results are not a function but a finite matrix of values.

The later issue is probably the easiest to solve. It means that we should fix a number of point N larger than 2^{J_n} and approximate the estimated copula density at step (7) on the induced grid $(i/N, j/N)$. We can not compute exactly the value on the grid as the scaling function is not always known explicitly. Nevertheless a very good approximation can be computed on this grid and we neglect the effect of this approximation. From the numerical point of view, this implies that the estimation error can be computed only on this grid and thus that the norms appearing in the numerical results (see Table 1, Table 2, Table 4 and Table 3) are empirical norms on this grid. In this paper, we choose $N = 4 * 2^{J_n}$. Note that step (3) also require an evaluation of the scaling function and thus is replaced by an approximation.

The former issue, the boundaries effect, is the key issue here. Indeed, for most

copula densities, the interesting behavior arises in the corners which are the most difficult part to handle numerically. The classical construction of the wavelet yields a basis over \mathbb{R}^2 while we only have samples on $[0, 1]^d$.

- A first choice is to consider the function of $[0, 1]^d$ to be estimated as a function of \mathbb{R}^2 which is null outside $[0, 1]^d$. This choice is called **zero padding** as it impose the value 0 outside $[0, 1]^d$.
- A second choice is to suppose that we observe the restriction on $[0, 1]^d$ of a 1-periodic function, this is equivalent to work in the classical periodic wavelet setting. This choice called **periodization** is very efficient if the function is really periodic.
- We propose also to modify the periodization and assume that we observe the restriction over $[0, 1]^d$ of a even 2-periodic function. As we introduce a symmetrization over the existing borders, we call this method **symmetrization**. It avoids the introduction of discontinuities along the border. Note that nevertheless this symmetrization introduces a discontinuities in the derivatives on the boundaries.

Once this extension has been performed on the sample, one can apply the classical wavelet transform. The resulting estimated copula density is the restriction to $[0, 1]^d$ of the estimated function.

The wavelet thresholding methods in a basis suffer from a gridding effect: we can often recognize a single wavelet in the estimated signal. To overcome this effect, we propose to use the **cycle spinning** trick proposed by Donoho and Johnstone. To add some translation invariance to the estimation process, we estimate the copula density with a collection of basis obtained by a simple translation from a single one and to average the resulting estimate. In our numerical experiments, we have performed this operation with 25 different translation parameters and observed that it has always ameliorate our estimate.

6.2 Simulation

We focus on the usual parametrical families of copulas: the FGM, the Gaussian, the Student, the Clayton, the Frank and the Gumbel families. We give results for both values of n (the number of data): $n = 500$ is very small for bidimensional problems and $n = 2000$ is usual in nonparametric estimation.

We test both methods of estimation and three ways to solve the boundaries problems. We simulated data with the same copula, the first margin being exponential with parameter 4 and the second margin being standard gaussian. Obviously, because of our algorithm, the results are exactly the same when we change the laws of the margins.

To evaluate the quality of our results, we consider three empirical loss functions deriving from the L_1 norm, L_2 norm and L_∞ norm:

$$E_q = \|\tilde{c} - c_0\|_{N,q} \quad \text{for } q = 1, 2, \infty$$

where c_0 is the "true" copula density and $N \times N$ is the number of points of the grid (see the previous part). Table 4 and Table 3 summarize the estimation relative errors

$$RE^q = \frac{\|\tilde{c} - c_0\|_{N,q}}{\|c_0\|_{N,q}} \quad \text{for } q = 1, 2, \infty.$$

These relative errors are computed with 100 repetitions of the experiment. The associated standard deviation is also given (in parentheses).

Table 1 and Table 2 show that the zero-padding method and the periodization method give similar results and lead to errors which are generally much larger than the symmetric periodization which is the best method to solve the boundaries effects. This remark is valid for $n = 500, 2000$. Although the zero-padding method is the default method with the Matlab Wavelet Toolbox, it suffers from a severe drawback: it introduces a strong discontinuities along the borders of $[0, 1]^d$. The periodization method suffers from the same drawback than the zero-padding method as soon as the function is not really periodic. Figure 1 emphasizes the superiority of the symmetric periodization method in the case where the unknown copula density is a normal copula. While the copula estimated with symmetric extension remains close from the shape of the true copula except for the resolution issue, this is not the case for the two other estimated copulas: in the periodized version, the height of the extreme peaks is reduced and two spurious peaks corresponding to the periodization of the real peaks appear. The zero padded version shows here only the reduced height artifact.

Tables 3 and 4 display the empirical L_1 , L_2 and L_∞ estimation error for the symmetric extension for respectively $n = 500$ and $n = 2000$. Globally, they show that the best results are obtained for the L_2 norm for which the method has been designed. The second best results are obtained for the L_1 norm because a bound on the L_2 norm implies a bound on the L_1 norm. The estimation in L_∞ is much harder as it is not a consequence of the estimation in L_2 and can be considered as a challenge for such a method.

One can also observed that the behavior largely depends on the copula itself. This is coherent with the theory that states that the more "regular" the copula is the more efficient the estimator will be. The copulas that are the least well estimated (Normal with parameter .9, Student with parameter .5 and Gumbel with parameter 8.33) are the most "irregular" ones. They are very "peaky" for the last two and almost singular along the first axis for the first one. They are therefore not enough regular to be estimated correctly by the proposed method.

A final remark should be given on the difficulty to evaluate such these errors. Whereas the L_1 norm is finite equal to 1 for all true copula, the L_2 and L_∞ norm can be very large even infinite because of their peaks. This is not an issue from the numerical point of view as we are restricted to a grid of step $1/N$ on which one can ensure the finiteness of the copula. Nevertheless the induce “empirical” norm can be substantially different from the integrated norm. Thus the error for $n = 500$ to $n = 2000$ are not strictly equivalent as the function can be much more complex for the resolution induced by $n = 2000$ than for $n = 500$.

6.3 Real data applications

We apply the thresholding methods on various financial series to identify the behavior of the dependance (or non dependance). All our data correspond to daily closing market quotations and are from 01/07/1987 to 31/01/2007. As usual, we consider the log–return of the data. Note that the data of each samples are not necessary independent but we apply our procedures as there were. We first propose estimators of the bivariate copula density associated with two financial series using the adaptive thresholding procedures (see Figure 2– Figure 5). Next, the nonparametrical estimator denoted \hat{c} is used as a benchmark and we derive an new estimator by minimization of the error between the benchmark and a given parametrical family of copula densities. As previously, we focus on copulas which belong on the Gaussian, Student, Gumbel, Clayton or Frank families. More precisely, we consider the following parametric classes of copulas

$$\begin{aligned}\mathcal{C}_1 &= \{c \in \mathcal{N}_\theta, \quad \theta = [-0.99 : 0.01 : 0.99]\} \\ \mathcal{C}_2 &= \{c \in \mathcal{T}_\theta, \quad \theta = [-0.99 : 0.01 : 0.99, 1 : 1 : 100]\} \\ \mathcal{C}_3 &= \{c \in \mathcal{G}_\theta, \quad \theta = [1 : 0.01 : 2]\} \\ \mathcal{C}_4 &= \{c \in \mathcal{C}_\theta, \quad \theta = [0 : 0.01 : 2]\} \\ \mathcal{C}_5 &= \{c \in \mathcal{F}_\theta, \quad \theta = [-2 : 0.01 : 2]\}\end{aligned}$$

We consider three distances

$$E_q(\theta, p) = \|\hat{c} - c_\theta\|_{N,q} \quad \text{for } q = 1, 2, \infty$$

where $c_\theta \in \mathcal{C}_p, p = 1, \dots, 5$. We propose to estimate the parameter θ for each class \mathcal{C}_p of copula densities as follows

$$\hat{\theta}_p^q = \arg \min_{\theta} E_q(\theta, p)$$

which appears to be the best estimator of θ under the constraint that the copula c belongs to a fixed parametrical family. We derive estimators of c among all the candidates $\{c_{\hat{\theta}_p^q}, p = 1, \dots, 5\}$ for each contrast $q = 1, 2, \infty$. Table 5– Table 8 give

- the parameter $\tilde{\theta}^q$ for $q = 1, 2, \infty$ defined by

$$\tilde{\theta}^q = \arg \min_{p=1,\dots,5} \left(\arg \min_{\theta} E_q(\theta, p) \right),$$

- the parametric family $\mathcal{C}_{\hat{p}}$ corresponding to the smallest error ,
- the associated relative errors defined by

$$RE^q(\tilde{\theta}^q) = 100 \frac{\|\hat{c} - c_{\tilde{\theta}^q}\|_{N,q}}{\|c_{\tilde{\theta}^q}\|_{N,q}}$$

where c is in $\mathcal{C}_{\hat{p}}$.

We test a lot of financial series and we select four revealing examples: we never observe that the Clayton family or the Gumbel family contains the best choice to model the unknown copula; the families used are always the Student class or the Frank class.

First, we observe that the results are very good since any computed relative error is small (in the worst case $RE^q \leq 20\%$). The results are quite similar for both thresholding methods when the unknown copula density does not present high picks (but the last one: DowJones versus Ftse100uk). In a theoretical point of view, we prefer the block thresholding method because the estimator are smoother. See by instance the case of the copula between Brent and ExxonMobil where the picks appearing in Figure 3 (on the right) are not pleasant (even if their ranges are not so large). Moreover, the relative error computed with the parametrical density which is the best one among all the possible parametrical copula densities is generally smallest for the block thresholding method.

Notice that the choice of the contrast is crucial to estimate the parameter θ : there are significative differences between $\hat{\theta}_1^p, \hat{\theta}_2^p, \hat{\theta}_\infty^p$. This is usual in density estimation. We prefer to measure the loss due to the estimator with the L_1 –norm because this norm is very adapted to the human eye and then the graphical results are the best. The quadratic loss is frequently used because the graphical results are easier to obtain but our opinion is that this norm does not emphasize enough differences between the densities. See by instance the very small relative errors computed with the L_2 contrast. The L_∞ norm has the opposite behavior: it accentuates every difference. It could be a drawback when the local thresholding method is considered and when too many details are kept (see again the case of the copula of the couple Brent/ExxonMobil).

Nevertheless, the choice of the best family do not depend on the choice of the contrast: it is fundamental because each type of parametrical family is linked to a specific behavior of the dependance and then the practitioner asks for indications about the copula type. The study of the copula Cac versus Brent allows to decide that both series are independent. Observe that there are small problems on the borders (but notice the very small scaling) although our methodology is made to remove this artefact. We think that a usual linear kernel method becomes disastrous when the copula density comes near of the uniform. The copula densities DowJones versus Oncedor and Brent versus ExxonMobil are both Frank copulas but with opposite behaviors. It seems natural that the series Brent and ExxonMobil are dependent and varying in the same sense. Oncedor (gold) is an hedge when the stock market collapses which could explain the negative dependance between Oncedor and the financial indices (we observe the same kind of dependance for others indices like Fste100uk, Cac ...). The more delicate case is for the copula DowJones versus Fste100uk because the picks are more accentuated. In this case, the local thresholding method produces a nice Figure.

In conclusion, we present here an estimation method which is popular among the practitioners: first, the nonparametrical estimator could be see as a benchmark to decide graphically if the unknown copula density looks like a copula density belonging to a well known parametrical family. In this case, the parameter is estimated with plug-in methods using the benchmark. We do not study here the properties of such an estimator or the goodness-of-fit test problem. For a statistic test procedure, we refer to Gayraud and Tribouley (2008).

7 Proofs

We first state the propositions needed to establish the main results. Next, we prove Theorem 4 in two steps by proving both implications. Last, we establish Proposition 1 and Corollary 5.1.

From now on we denote K any constant that may change from one line to another, which does not depend on j , k and n but which depends on the wavelet and on $\|c\|_\infty$ and $\|c\|_2$.

7.1 Preliminaries

These preliminary results concern the estimation of the wavelet coefficients and the scaling coefficients (denoted $c_{j,k}^{\epsilon_0}$ with $\epsilon_0 = (0, \dots, 0)$ to unify the notation). Proposition 3 announces that the accuracy of the estimation is as

sharp as if the direct observations were available.

Proposition 2 *Assume that the copula density belongs to L_∞ and let $\delta > 0$. There exists a constant $K > 0$ such that for any j such that $2^j \leq 2 \left(\frac{n}{\log(n)}\right)^{1/d}$, and for any (k, ϵ)*

$$\mathbb{P} \left(|\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon}| > \lambda_n \right) \leq K n^{-\delta} \quad (9)$$

$$\mathbb{P} \left(\sum_k (\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon})^2 > L^d 2^{dj} \lambda_n^2 \right) \leq K n^{1-\delta} (\log(n))^{-1} \quad (10)$$

as soon as κ is chosen large enough.

It is clear that (10) is a direct consequence of (9). Proof of (9) is rejected to the Appendix. Note that from (9) we immediately deduce

Proposition 3 *Under the same assumptions as in Proposition 2 on j and c , there exists a constant $K > 0$ such that for any (k, ϵ)*

$$E \left[\left(\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon} \right)^2 \right] \leq K \frac{\log(n)}{n}.$$

7.2 Proof of Theorem 4

First, we prove the result for the linear estimator. Secondly, we prove the result for the local thresholding method. We do not prove the result for the global thresholding method since the technics are the same except that the required large deviation inequality established in Proposition 2 is given by (10) instead of (9).

7.2.1 Proof of Equivalence (6)

On the one hand, let c be a copula density function belonging to L_∞ satisfying for any n ,

$$\mathbb{E} \|\widetilde{c_L} - c\|_2^2 \leq K \left(\frac{\log(n)}{n} \right)^{\frac{2s}{2s+d}} \quad (11)$$

for some constant $K > 0$. Let us prove that c also belongs to the space $\mathcal{B}_{2\infty}^s$. Let us recall that the smoothing index used for the linear procedure is j_n^* satisfying $2^{1-j_n^*} > (n^{-1} \log(n))^{1/(2s+d)}$. Since

$$\mathbb{E}\|\widetilde{c}_L - c\|_2^2 = \mathbb{E}\|\widetilde{c}_L - \sum_k c_{j_n^*, k} \phi_{j_n^*, k}\|_2^2 + \|\sum_{j \geq j_n^*} \sum_{k, \epsilon} c_{j, k}^\epsilon \psi_{j, k}^\epsilon\|_2^2,$$

and following the assumption (11), we get

$$\sum_{j \geq j_n^*} \sum_{k, \epsilon} (c_{j, k}^\epsilon)^2 \leq \mathbb{E}\|\widetilde{c}_L - c\|_2^2 \leq K (2^{-2j_n^*})^s$$

which is the announced result. On the other hand, let us suppose that $c \in \mathcal{B}_{2\infty}^s$. Then, using the same technics as in Genest et al. [9], we prove that

$$\mathbb{E}\|\widetilde{c}_L - c\|_2^2 \leq K \left(\frac{\log(n)}{n} \right)^{\frac{2s}{2s+d}}$$

which ends the proof. The proof in Genest et al. [9] is given in the case $d = 2$ and need some sharp control on the estimated coefficients because an optimal result is established (there is no logarithmic term in the rate).

7.2.2 Proof of Equivalence (7) (first step: \implies)

When direct observations $(F_1(X_i^1), \dots, F_d(X_i^d))$ are available, we use the estimator \widehat{c}_{HL} built in the same way as \widetilde{c}_{HL} but with the sequence of coefficients $\widehat{c}_{j, k}^\epsilon$ defined in (1) and the threshold $\lambda_n/2$ instead of λ_n . Let us take j_n, J_n positive integers and $\lambda_n > 0$. Since we get

$$E\|\widehat{c}_{HL} - c\|_2^2 \leq 2E\|\widehat{c}_{HL} - \widetilde{c}_{HL}\|_2^2 + 2E\|\widetilde{c}_{HL} - c\|_2^2$$

we have then to study the error term due to the fact that we use pseudo observations

$$\begin{aligned}
T &= E \|\widehat{c_{HL}} - \widehat{c_{HL}}\|_2^2 \\
&= E \left[\sum_k (\widetilde{c_{j_n k}^{\epsilon_0}} - \widehat{c_{j_n k}^{\epsilon_0}})^2 \right] + E \left[\sum_{j_n} \sum_{k, \epsilon} (\widetilde{c_{j, k}^{\epsilon}} - \widehat{c_{j, k}^{\epsilon}})^2 \mathbf{1}\{|\widetilde{c_{j, k}^{\epsilon}}| > \lambda_n\} \mathbf{1}\{|\widehat{c_{j, k}^{\epsilon}}| > \frac{\lambda_n}{2}\} \right] \\
&\quad + E \left[\sum_{j_n} \sum_{k, \epsilon} (\widehat{c_{j, k}^{\epsilon}})^2 \mathbf{1}\{|\widetilde{c_{j, k}^{\epsilon}}| \leq \lambda_n\} \mathbf{1}\{|\widehat{c_{j, k}^{\epsilon}}| > \frac{\lambda_n}{2}\} \right] \\
&\quad + E \left[\sum_{j_n} \sum_{k, \epsilon} (\widetilde{c_{j, k}^{\epsilon}})^2 \mathbf{1}\{|\widetilde{c_{j, k}^{\epsilon}}| > \lambda_n\} \mathbf{1}\{|\widehat{c_{j, k}^{\epsilon}}| \leq \frac{\lambda_n}{2}\} \right] \\
&= T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

Using Proposition 3, we have

$$T_1 \leq K \frac{\log(n)}{n} 2^{dj_n} \leq K \frac{(\log(n))^2}{n}. \quad (12)$$

For the study of T_2 , we separate the cases where the wavelet coefficients are larger or smaller than the thresholding level $\lambda_n/4$. By Cauchy-Schwartz Inequality, we have

$$\begin{aligned}
T_2 &= E \left[\sum_{j_n} \sum_{k, \epsilon} (\widetilde{c_{j, k}^{\epsilon}} - \widehat{c_{j, k}^{\epsilon}})^2 \mathbf{1}\{|\widetilde{c_{j, k}^{\epsilon}}| > \lambda_n\} \mathbf{1}\{|\widehat{c_{j, k}^{\epsilon}}| > \frac{\lambda_n}{2}\} \left(\mathbf{1}\{|c_{j, k}^{\epsilon}| \leq \frac{\lambda_n}{4}\} + \mathbf{1}\{|c_{j, k}^{\epsilon}| > \frac{\lambda_n}{4}\} \right) \right] \\
&\leq \sum_{j_n} \sum_{k, \epsilon} \left[E(\widetilde{c_{j, k}^{\epsilon}} - \widehat{c_{j, k}^{\epsilon}})^4 \right]^{1/2} \left[\mathbb{P} \left(|\widehat{c_{j, k}^{\epsilon}} - c_{j, k}^{\epsilon}| > \frac{\lambda_n}{4} \right) \right]^{1/2} \\
&\quad + \sum_{j_n} \sum_{k, \epsilon} E(\widetilde{c_{j, k}^{\epsilon}} - \widehat{c_{j, k}^{\epsilon}})^2 \mathbf{1}\{|c_{j, k}^{\epsilon}| > \frac{\lambda_n}{4}\}.
\end{aligned}$$

Observe that, for any j, k, ϵ , we have

$$|\widetilde{c_{j, k}^{\epsilon}}| \vee |\widehat{c_{j, k}^{\epsilon}}| \leq 2^{jd/2} (\|\psi\|_{\infty}^d \vee \|\phi\|_{\infty}^d). \quad (13)$$

For any $\delta > 0$, we use now the standard Bernstein Inequality to obtain

$$\mathbb{P} \left(|\widehat{c_{j, k}^{\epsilon}} - c_{j, k}^{\epsilon}| > \frac{\lambda_n}{4} \right) \leq K n^{-\delta} \quad (14)$$

which is valid for a choice of κ large enough. Let us now fix r in $]0, 2[$. Applying Proposition 3 and using (13), it follows

$$\begin{aligned}
T_2 &\leq K \sum_{j_n} \sum_{k,\epsilon} 2^{jd} \left[\mathbb{P} \left(|\widehat{c_{j,k}^\epsilon} - c_{j,k}^\epsilon| > \frac{\lambda_n}{4} \right) \right]^{1/2} + \sum_{j_n} \sum_{k,\epsilon} E(\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon})^2 \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \\
&\leq K \left(2^{2dJ_n} n^{-\delta/2} + u_n \left[\left(\frac{\lambda_n}{4} \right)^r \sum_{j_n} \sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \right] \right)
\end{aligned}$$

for

$$u_n = \left(\frac{\lambda_n}{4} \right)^{-r} \frac{\log(n)}{n}.$$

Similarly, we have

$$\begin{aligned}
T_3 &\leq E \left[\sum_{j_n} \sum_{k,\epsilon} (\widehat{c_{j,k}^\epsilon})^2 \mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| \leq \lambda_n\} \mathbf{1}\{|\widehat{c_{j,k}^\epsilon}| > \frac{\lambda_n}{2}\} \left(\mathbf{1}\{|c_{j,k}^\epsilon| \leq \frac{\lambda_n}{4}\} + \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \right) \right] \\
&\leq E \left[\sum_{j_n} \sum_{k,\epsilon} (\widehat{c_{j,k}^\epsilon})^2 \mathbf{1}\{|\widehat{c_{j,k}^\epsilon}| > \frac{\lambda_n}{2}\} \mathbf{1}\{|c_{j,k}^\epsilon| \leq \frac{\lambda_n}{4}\} \right] \\
&\quad + E \left[\sum_{j_n} \sum_{k,\epsilon} (\widehat{c_{j,k}^\epsilon})^2 \mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| \leq \lambda_n\} \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \right] \\
&\leq K \sum_{j_n} \sum_{k,\epsilon} 2^{dj} \mathbb{P} \left(|\widehat{c_{j,k}^\epsilon} - c_{j,k}^\epsilon| > \frac{\lambda_n}{4} \right) + \left(\frac{\lambda_n}{4} \right)^r \sum_{j_n} \sum_{k,\epsilon} v_n \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\}
\end{aligned}$$

for

$$\begin{aligned}
v_n &= 2 \left(\frac{\lambda_n}{4} \right)^{-r} \left[E(\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon})^2 + E(\widetilde{c_{j,k}^\epsilon})^2 \mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| \leq \lambda_n\} \right] \\
&\leq 2 \left(K u_n + 4^r \lambda_n^{2-r} \right)
\end{aligned}$$

implying that

$$T_3 \leq K \left(\frac{2^{2dJ_n}}{n^\delta} + (u_n + \lambda_n^{2-r}) \left[\left(\frac{\lambda_n}{4} \right)^r \sum_{j_n} \sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \right] \right).$$

Using (13) and Proposition 2, we get

$$T_4 \leq K \sum_{j_n} \sum_{k,\epsilon} 2^{dj} \mathbb{P} \left(|\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon}| > \frac{\lambda_n}{2} \right) \leq K 2^{2dJ_n} n^{-\delta}.$$

Combining the bounds of T_1, T_2, T_3, T_4 and choosing j_n, J_n as indicated in Theorem 2, we get for $\delta \geq 6$

$$E\|\widehat{c_{HL}} - c\|_2^2 \leq 2 E\|\widehat{c_{HL}} - c\|_2^2 + K \rho_n$$

where

$$\rho_n = \frac{(\log(n))^2}{n} + \left(\frac{\log n}{n}\right)^{1-\frac{r}{2}} \left(\frac{\lambda_n}{4}\right)^r \sum_{j_n} \sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} + \frac{1}{n(\log(n))^2}.$$

On the one hand, let us suppose that c belongs to the weak Besov $\mathcal{W}_L(\frac{2d}{2s+d})$ which means that (for $r := 2d/(2s+d)$)

$$\left(\frac{\lambda_n}{4}\right)^r \sum_{j_n} \sum_{k,\epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \frac{\lambda_n}{4}\} \leq K.$$

It follows that

$$\rho_n \leq K \left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}.$$

Using the standard result given in Theorem 3 when direct observations are available, we also have

$$E\|\widehat{c_{HL}} - c\|_2^2 \leq K \left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}}$$

as soon as $c \in \mathcal{W}_L(\frac{2d}{2s+d}) \cap \mathcal{B}_{2\infty}^s$. This ends the proof of the first part of (7) of Theorem 4.

7.2.3 Proof of Equivalence (7) (second step: \Leftarrow)

Suppose that there exists M such that for any n , $E\|\widehat{c_{HL}} - c\|_2^2 \leq M (n^{-1} \log(n))^{\frac{2s}{2s+d}}$. Since

$$\sum_{j > J_n} \sum_{k,\epsilon} (c_{j,k}^\epsilon)^2 \leq E\|\widehat{c_{HL}} - c\|_2^2,$$

and choosing J_n as indicated in Theorem 2, we obtain

$$\sum_{j > J_n} \sum_{k,\epsilon} (c_{j,k}^\epsilon)^2 \leq M \left(\frac{\log(n)}{n}\right)^{\frac{2s}{2s+d}} \leq M (2^{d(1-J_n)})^{\frac{2s}{2s+d}} \leq K (2^{-2J_n})^{\frac{ds}{2s+d}}.$$

Using Definition 1 of the strong Besov spaces, we deduce that c belongs necessarily to $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$. Let us now study the sum of the square of the details when the details coefficients are small

$$\begin{aligned} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \frac{\lambda_n}{2}\} &= \left[\sum_{j < j_n} + \sum_{j=j_n}^{J_n} + \sum_{j > J_n} \right] \left[\sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \lambda_n/2\} \right] \\ &\leq H_1 + H_2 + H_3. \end{aligned}$$

Since we have already proved that $c \in \mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$ and taking λ_n as indicated in Theorem 2, we deduce

$$H_3 \leq \sum_{j > J_n} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \leq K 2^{-2J_n \frac{ds}{2s+d}} \leq K \left(\frac{\lambda_n}{2} \right)^{\frac{4s}{2s+d}}.$$

Taking j_n as in Theorem 2, we get

$$H_1 \leq K \sum_{j < j_n} 2^{dj} \left(\frac{\lambda_n}{2} \right)^2 \leq K \log(n) \left(\frac{\lambda_n}{2} \right)^2 \leq K \left(\frac{\lambda_n}{2} \right)^{\frac{4s}{2s+d}}.$$

Observe that

$$H_2 = E \left[\sum_{j_n}^{J_n} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \frac{\lambda_n}{2}\} \left(\mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| \leq \lambda_n\} + \mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| > \lambda_n\} \right) \right].$$

Remembering that

$$E \left[\sum_{j_n}^{J_n} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|\widetilde{c_{j,k}^\epsilon}| \leq \lambda_n\} \right] \leq E \|\widetilde{c_{HL}} - c\|_2^2$$

and using Proposition 2 and (14), we get

$$\begin{aligned} H_2 &\leq E \|\widetilde{c_{HL}} - c\|_2^2 + \sum_{j_n}^{J_n} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbb{P}(|\widetilde{c_{j,k}^\epsilon} - \widehat{c_{j,k}^\epsilon}| > \frac{\lambda_n}{4}) \\ &\quad + \sum_{j_n}^{J_n} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbb{P}(|\widehat{c_{j,k}^\epsilon} - c_{j,k}^\epsilon| > \frac{\lambda_n}{4}) \\ &\leq M \left(\frac{\log(n)}{n} \right)^{\frac{2s}{2s+d}} + K \|c\|_2^2 n^{-\delta} \leq K \left(\frac{\lambda_n}{2} \right)^{\frac{4s}{2s+d}} \end{aligned}$$

as soon as δ is larger than 1. Combining using Definition 2 of the local weak Besov space, we conclude that $c \in \mathcal{W}_L(r)$ with r such that $2-r = 4s/(2s+d)$. Hence, we end the proof of the indirect direction of (7).

7.3 Proofs of Proposition 1 and Corollary 5.1

The proof of the **large inclusion** given in Proposition 1 follows immediately from the definitions of the functional spaces. Denote $c_{j,k}^\epsilon$ the sequence of wavelet coefficients of a function c . Since we have

$$\begin{aligned}
& \sup_{0 < \lambda \leq 1} \lambda^{r-2} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \lambda\} \\
&= \sup_{0 < \lambda \leq 1} \lambda^{r-2} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\{|c_{j,k}^\epsilon| \leq \lambda\} \left[\mathbf{1}\left\{\sum_k (c_{j,k}^\epsilon)^2 \leq 2^{dj} \lambda^2\right\} \right. \\
&\quad \left. + \mathbf{1}\left\{\sum_k (c_{j,k}^\epsilon)^2 > 2^{dj} \lambda^2\right\} \right] \\
&\leq \sup_{0 < \lambda \leq 1} \lambda^{r-2} \sum_{j \geq 0} \sum_{k, \epsilon} (c_{j,k}^\epsilon)^2 \mathbf{1}\left\{\sum_k (c_{j,k}^\epsilon)^2 \leq 2^{dj} \lambda^2\right\} \\
&\quad + K \sup_{0 < \lambda \leq 1} \lambda^r \sum_{j \geq 0} 2^{dj} \sum_{\epsilon} \mathbf{1}\left\{\sum_k (c_{j,k}^\epsilon)^2 > 2^{dj} \lambda^2\right\},
\end{aligned}$$

it follows from Definition 3 that

$$c \in \mathcal{W}_G(r) \Rightarrow c \in \mathcal{W}_L(r).$$

To establish the **strict inclusions**, we build a sparse function belonging to $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}} \cap \mathcal{W}_L(\frac{2d}{2s+d})$ but not to $\mathcal{W}_G(\frac{2d}{2s+d})$. Let us choose a real number α such that $\frac{d}{2} \leq \alpha < s + \frac{d}{2}$. Let us consider a function c with the sparse sequence $c_{j,k}^\epsilon$ such that at each level $j \in \mathbb{N}$ and at each $\epsilon \in S_d$, only $\lfloor 2^{\frac{2d\alpha}{2s+d}j} \rfloor$ wavelet coefficients take the value $(2^d - 1)^{-1} 2^{-\alpha j}$. The others are equal to 0. For all $0 < \lambda \leq 1$, let j_λ be such that $2^{j_\lambda} = ((2^d - 1)\lambda)^{-\frac{1}{\alpha}}$. We get

$$\begin{aligned}
\sum_{j \geq 0} \sum_{k, \epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \lambda\} &= \sum_{j < j_\lambda} \sum_{k, \epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \lambda\} \\
&\leq K 2^{\frac{2d\alpha}{2s+d}j_\lambda} \leq K \lambda^{-\frac{2d}{2s+d}}.
\end{aligned}$$

implying that

$$\sup_{0 < \lambda \leq 1} \lambda^{\frac{2d}{2s+d}} \sum_{j \geq 0} \sum_{k, \epsilon} \mathbf{1}\{|c_{j,k}^\epsilon| > \lambda\} < \infty,$$

and the function c belongs to the local weak Besov space $\mathcal{W}_L(\frac{2d}{2s+d})$. Next, put $\alpha' = (4\alpha s + 2sd + d^2)/(2(2s + d))$ and observe that $\alpha' < s + d/2$ since $\alpha < s + d/2$. For all $0 < \lambda \leq 1$ let j_λ^* be such that $2^{j_\lambda^*} = ((2^d - 1)\lambda)^{-\frac{1}{\alpha'}}$. We get

$$\begin{aligned} \sum_{j \geq 0} 2^{dj} \sum_{\epsilon} \mathbf{1} \{ \sum_k (c_{j,k}^{\epsilon})^2 > 2^{dj} \lambda^2 \} &\geq (2^d - 1) \sum_{j < j_{\lambda}^*} 2^{dj} \\ &\geq 2^{dj_{\lambda}^* - 1} \geq K \lambda^{-\frac{d}{\alpha'}}. \end{aligned}$$

implying

$$\sup_{0 < \lambda \leq 1} \lambda^{\frac{2d}{2s+d}} \sum_{j \geq 0} 2^{dj} \sum_{\epsilon} \mathbf{1} \{ \sum_k (c_{j,k}^{\epsilon})^2 > 2^{dj} \lambda^2 \} = \infty.$$

It follows that the function c does not belong to the global weak Besov $\mathcal{W}_G(\frac{2d}{2s+d})$ which ends the proof of Proposition 1. Notice that the function c belongs to the strong Besov space $\mathcal{B}_{2\infty}^{\frac{ds}{2s+d}}$ because for all (j, ϵ)

$$\sum_{k, \epsilon} (c_{j,k}^{\epsilon})^2 \leq 2^{\frac{2d\alpha}{2s+d}j} 2^{-2\alpha j} \leq 2^{-\frac{2ds}{2s+d}j}$$

implying that

$$\sup_{J \geq 0} 2^{\frac{2ds}{2s+d}J} \sum_{j \geq J} \sum_{k, \epsilon} (c_{j,k}^{\epsilon})^2 < \infty.$$

Corollary 5.1 is also proved too.

8 Appendix

This section aims at proving (9) of Proposition 2. In the sequel we fix the indices j and $k = (k_1, \dots, k_d)$ and take without loss of generality $\epsilon = 2^d - 1$. For any $i = 1, \dots, n$ (the observation index) and any $m = 1, \dots, d$ (the coordinate index), let us introduce the following notations

$$\begin{aligned} \Delta(X_i^m) &= \widehat{F}_m(X_i^m) - F_m(X_i^m), \\ \xi_j(X_i^m) &= \psi_{j,k_m}(\widehat{F}_m(X_i^m)) - \psi_{j,k_m}(F_m(X_i^m)), \\ N_j(m) &= \# \{i \in \{1, \dots, n\}; \xi_j(X_i^m) \neq 0\}, \end{aligned}$$

as univariate quantities and

$$\begin{aligned} \xi_j(X_i^1, \dots, X_i^m) &= \psi_{j,k}^{\epsilon}(\widehat{F}_1(X_i^1), \dots, \widehat{F}_d(X_i^d)) - \psi_{j,k}^{\epsilon}(F_1(X_i^1), \dots, F_d(X_i^d)) \\ N_j &= \# \{i \in \{1, \dots, n\}; \xi_j(X_i^1, \dots, X_i^d) \neq 0\} \end{aligned}$$

as d -variate quantities. As previously remarked in Genest et al. [9] (for $d = 2$), from the definitions above we have

$$\begin{aligned}
\xi_j(X_i^1, \dots, X_i^d) &= \prod_{m=1}^d \xi_j(X_i^m) + \sum_{m_1=1}^d \left[\psi_{j,k_{m_1}}^\epsilon(F_{m_1}(X_i^{m_1})) \prod_{\substack{m=1 \\ m \neq m_1}}^d \xi_j(X_i^m) \right] \\
&+ \sum_{\substack{m_1, m_2=1 \\ m_1 \neq m_2}}^d \left[\psi_{j,k_{m_1}}^\epsilon(F_{m_1}(X_i^{m_1})) \psi_{j,k_{m_2}}^\epsilon(F_{m_2}(X_i^{m_2})) \prod_{\substack{m=1 \\ m \neq m_1, m_2}}^d \xi_j(X_i^m) \right] \\
&+ \dots + \sum_{m_1=1}^d \left[\xi_j(X_i^{m_1}) \prod_{\substack{m=1 \\ m \neq m_1}}^d \psi_{j,k_m}^\epsilon(F_m(X_i^m)) \right]. \tag{15}
\end{aligned}$$

In the sequel, for $m = 1, \dots, d$, we denote by $T_{m,j}(X_i)$ any term of the type

$$\left[\psi_{j,k_1}^\epsilon(F_1(X_i^1)) \times \dots \times \psi_{j,k_{d-m}}^\epsilon(F_{d-m}(X_i^{d-m})) \right] \left[\xi_j(X_i^{d-m+1}) \times \dots \times \xi_j(X_i^d) \right]$$

i.e. such that there are exactly m factors $\xi_j(X_i)$ appearing in the product. The cardinality of such terms $T_{m,j}(X_i)$ is equal to $C_d^m = \frac{d!}{m!(d-m)!}$. Observe that the number of terms in (15) is $2^d - 1$. It is fundamental to notice that there is no term $T_{0,j}(X_i) = \prod_{m=1, \dots, d} \psi_{j,k_m}^\epsilon(F_m(X_i^m))$.

8.1 Technical lemmas

We begin by giving technical lemmas.

Lemma 1 *There exists a universal constant K_0 such that for any $m \in \{1, \dots, d\}$*

$$\forall t > 0, \mathbb{P}(\max_{1 \leq i \leq n} |\Delta(X_i^m)| > t) \leq K_0 \exp(-2nt^2).$$

Lemma 1 is a consequence of the Dvoretzky-Kiefer-Wolfowitz inequality. For the interested reader, the detailed proof of this lemma is given in Autin et al. [3].

Lemma 2 *Let $\delta > 0$ and n be an integer such that $n \log(n) \geq 2(\delta^{-1} \vee 1)$. Then, there exists $K_1 > 0$ such that for any level j satisfying*

$$2^j \leq \frac{1}{3} \left(\frac{2n}{\delta \log(n)} \right)^{1/2},$$

and for any $m \in \{1, \dots, d\}$,

$$\mathbb{P}(N_j(m) > (L+3)n2^{-j}) \quad \vee \quad \mathbb{P}(N_j > d(L+3)n2^{-j}) \leq K_1 n^{-\delta}.$$

The proof of Lemma 2 can be found in Autin et al. [3]

Lemma 3 *Let us assume that c belongs to L_∞ and let $(j, N) \in \mathbb{N}^2$. For all $1 \leq p \leq q \leq d$, for all subsets \mathcal{S}_p and \mathcal{S}_{q-p} of $\{1, \dots, d\}$ with cardinalities equal to p and $q-p$ having no common component, let us put for $i = 1 \dots, n$,*

$$Z_i(\mathcal{S}_p, \mathcal{S}_{q-p}) = \prod_{m \in \mathcal{S}_p} \psi_{j,k_m}(F_m(X_i^m)) \prod_{m' \in \mathcal{S}_{q-p}} (\psi^{(1)})_{j,k_{m'}}(F_{m'}(X_i^{m'})), \quad (16)$$

with the following notation $\psi_{j,k}^{(1)}(\cdot) = 2^{j/2} \psi'(2^j \cdot - k)$.
For any $\mu \geq 2K_3 2^{-jq/2}$, we have

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N Z_i(\mathcal{S}_p, \mathcal{S}_{q-p}) \right| > \mu \right) \leq 2 \exp \left(-K_2 N \left(\mu^2 \wedge \mu 2^{1-jq/2} \right) \right)$$

where K_2, K_3 are constants such that

$$K_3 \geq (L+1)^q \|c\|_\infty \|\psi\|_\infty^p \|\psi'\|_\infty^{q-p}, \quad K_2 \leq \frac{1}{8} \|\psi\|_\infty^{-p} \|\psi'\|_\infty^{p-q} (K_3^{-1} \vee 6).$$

Lemma 3 is a direct application of the Bernstein Inequality with

$$|EZ_i(\mathcal{S}_p, \mathcal{S}_{q-p})|$$

$$\begin{aligned} &= \left| \int_{[0,1]^d} \prod_{m \in \mathcal{S}_p} \psi_{j,k_m}(u_m) \prod_{m' \in \mathcal{S}_{q-p}} \psi_{j,k_{m'}}^{(1)}(u_{m'}) c(u_1, \dots, u_d) du_1 \times \dots \times du_d \right| \\ &\leq K_3 2^{-jq/2} \end{aligned}$$

and in the same way,

$$\text{Var}(Z_i(\mathcal{S}_p, \mathcal{S}_{q-p})) \leq (L+1)^q \|c\|_\infty \|\psi\|_\infty^{2p} \|\psi'\|_\infty^{2(q-p)}$$

and

$$|Z_i(\mathcal{S}_p, \mathcal{S}_{q-p})| \leq \|\psi\|_\infty^p \|\psi'\|_\infty^{q-p} 2^{jq/2}.$$

8.2 Proof of Proposition 2

By Equality (15), we have for any $\lambda > 0$

$$\mathbb{P}\left(\left|\widehat{c_{j,k}^\epsilon} - \widetilde{c_{j,k}^\epsilon}\right| > \lambda\right) \leq \sum_{m=1}^d C_d^m L_m$$

for

$$L_m = \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n T_{m,j}(X_i)\right| > \frac{\lambda}{2^d - 1}\right).$$

Using a Taylor expansion, the following inequality holds for $i = 1, \dots, n$ and $m' = 1, \dots, d$

$$|\xi_j(X_i^{m'})| \leq 2^j |\Delta(X_i^{m'})| (\psi^{(1)})_{j,k_{m'}}(F_{m'}(X_i^{m'})) + 2^{\frac{3j}{2}-1} |\Delta(X_i^{m'})|^2 \|\psi'\|_\infty$$

implying that, for an associated \mathcal{S}_{d-m}

$$|T_{m,j}(X_i)| \leq \|\psi'\|_\infty^m \sum_{\substack{m'=0 \\ \mathcal{S}_{m-m'} \cap \mathcal{S}_{d-m'} = \emptyset}}^m 2^{j(m+m'/2)} \left(\max_{m'=1, \dots, m} |\Delta(X_i^{m'})| \right)^{m+m'} |Z_i(\mathcal{S}_{d-m}, \mathcal{S}_{m-m'})|.$$

For $m = 1, \dots, d$, let us introduce the following events

$$\mathcal{D}_{0,m} = \left\{ \max_{1 \leq i \leq n} |\Delta(X_i^m)| \leq \sqrt{\frac{\delta \log(n)}{2n}} \right\}, \mathcal{D}_{1,m} = \{N_j(m) \leq n_j = (L+3)n2^{-j}\},$$

$$\mathcal{D}_0 = \bigcap_{m=1}^d \mathcal{D}_{0,m}, \mathcal{D}_1 = \bigcap_{m=1}^d \mathcal{D}_{1,m}.$$

It follows that for any \mathcal{S}_p and any \mathcal{S}_{q-p}

$$\begin{aligned} L_m &\leq \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n T_{m,j}(X_i)\right| > \frac{\lambda}{2^d - 1}\right) \cap \mathcal{D}_0 \cap \mathcal{D}_1 + \mathbb{P}(\mathcal{D}_0^c) + \mathbb{P}(\mathcal{D}_1^c) \\ &\leq \sum_{m'=0}^m \mathbb{P}\left(\left|\frac{1}{n_j} \sum_{i=1}^{n_j} Z_i(\mathcal{S}_{d-m}, \mathcal{S}_{m-m'})\right| > \mu\right) + \mathbb{P}(\mathcal{D}_0^c) + \mathbb{P}(\mathcal{D}_1^c) \end{aligned}$$

where

$$\mu = 2^{-j(m+m'/2)} \left(\frac{2n}{\delta \log(n)} \right)^{\frac{m+m'}{2}} \frac{2^j \|\psi'\|_\infty^{-m} (L+3)^{-1} \lambda}{(2^d - 1)(m+1) C_m^{\lfloor m/2 \rfloor}}.$$

Fix $\kappa > 0$ and take $\lambda = \sqrt{\frac{\kappa \log(n)}{n}}$. Using Lemma 1 and Lemma 2, one gets

$$\mathbb{P}(\mathcal{D}_0^c) \vee \mathbb{P}(\mathcal{D}_1^c) \leq d(K_0 \vee K_1) n^{-\delta} \quad (17)$$

as soon as $2^j \leq \frac{1}{3} \left(\frac{2n}{\delta \log(n)} \right)^{1/2}$. Since $\mu \geq 2K_3 2^{-j(d-m')/2}$, we apply Lemma 3 and we obtain

$$L_m \leq 2 \sum_{m'=0}^m \exp \left[-K_2 2^{-j} n \left(\mu^2 \wedge \mu 2^{1-j(d-m')/2} \right) \right] + d(K_0 \vee K_1) n^{-\delta} \leq K n^{-\delta}$$

as soon as

$$\mu \geq \left(\frac{\delta}{K_2} \frac{2^j \log(n)}{n} \right)^{1/2} \vee \left(\frac{\delta}{K_2} \frac{2^{j(2+d-m')/2} \log(n)}{2n} \right). \quad (18)$$

Let us restrict ourselves to the case where:

$$2^j \leq \left(\frac{n}{\log n} \right)^{1/d}.$$

Assuming that n is large enough and that κ is chosen large enough, Condition (18) on μ is satisfied if, for any $m' = 0, \dots, m$

$$d \geq \frac{2m + m' - 1}{m + m'} \vee \frac{2m + d}{m + m' + 1}.$$

This condition is always satisfied since $d \geq 2$. We obtain the announced result.

References

- [1] Autin, F. (2006) Maxisets for density estimation on \mathbb{R} . (2006). *Math. Methods of Statist.*, n 2, 123-145.
- [2] Autin, F., Picard, D. and Rivoirard, V. (2006). Maxiset approach for Bayesian nonparametric estimation. *Math. Methods of Statist.*, vol 15, n 4, 349-373.
- [3] Autin, F., Le Pennec, E. and Tribouley, K. (2008). Thresholding methods to estimate the copula density. preprint on the web (<http://www.cmi.univ-mrs.fr/~autin/DONNEES/COPULAS>).

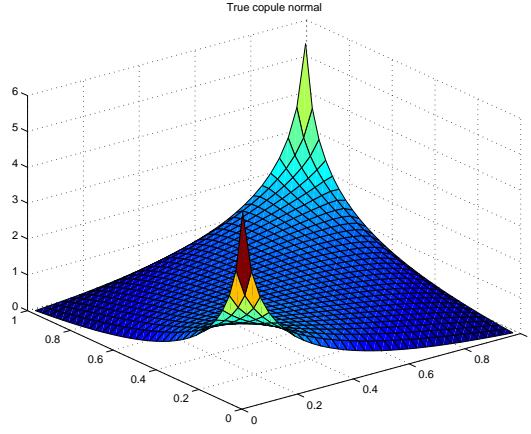
- [4] Cohen, A., De Vore, R., Kerkycharian, G. and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, **11**, 167-191.
- [5] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- [6] Donoho, D.L., Johnstone, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.*, **90**(432), 1200-1224.
- [7] Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D.(1996). Density estimation by wavelet thresholding. *Annals of Statistics*, **24**, 508-539.
- [8] Gayraud, G., and Tribouley, K. (2008). Good-fit-of test for the copula density. *Submitted paper*.
- [9] Genest, C., Masiello, E. and Tribouley, K. (2008). Estimating copula densities through wavelets. *Submitted paper*.
- [10] Hoffmann, M. and Lepski, O. V. (2002). Random rates in anisotropic regression. With discussions and a rejoinder by the authors. *Ann. Statist.*, **30**,(2), 325-396.
- [11] Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical estimation*. Springer-Verlag, New-York. Asymptotic theory, translated from the Russian by Samuel Kotz.
- [12] Kerkycharian, G., Picard, D. (1992). Density estimation in Besov spaces. *Statistics and Probability Letters*, **13**, 15-24.
- [13] Kerkycharian, G., Picard, D., Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli*, **2** 229-247.
- [14] Kerkycharian, G., Picard, D. (2001). Thresholding algorithms, maxisets and well concentrated bases. *Test*, **9** n 2, 283-344.
- [15] Kerkycharian, G., Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, **10** 1053-1105.
- [16] Meyer, Y. (1990). *Ondelettes et Opérateurs*, Hermann, Paris.
- [17] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229-231.

Correspondances:

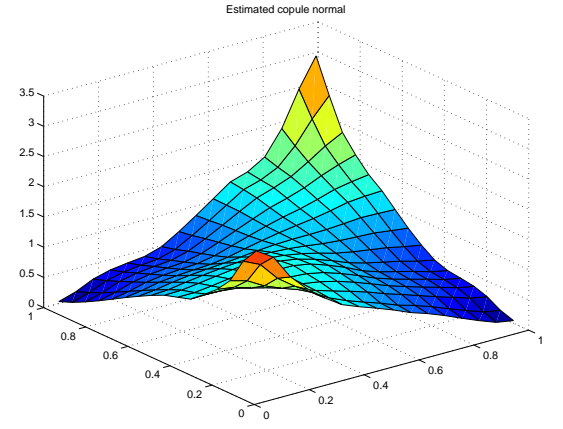
AUTIN Florent, LATP, université Aix-Marseille 1, Centre de Mathématiques et Informatique, 39 rue F. Joliot Curie, 13453 Marseille Cedex 13 (autin@cmi.univ-mrs.fr).

LEPENNEC Erwan, LPMA, université Paris 7, 175 rue du Chevaleret, 75013 Paris (lepenne@math.jussieu.fr).

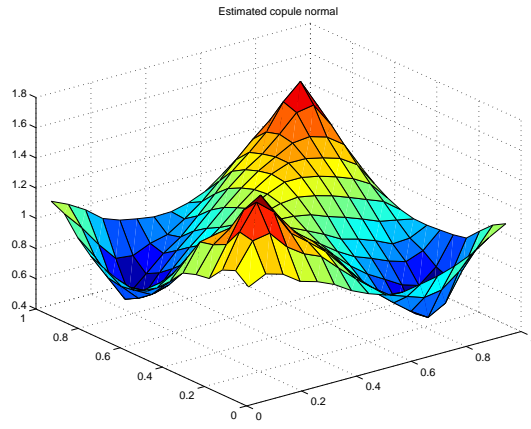
TRIBOULEY Karine, LPMA, 175 rue du Chevaleret, 75013 Paris and MODALX, Université Paris 10-Nanterre, 200 avenue de la République, 92001 Nanterre Cedex (ktriboul@u-paris10.fr).



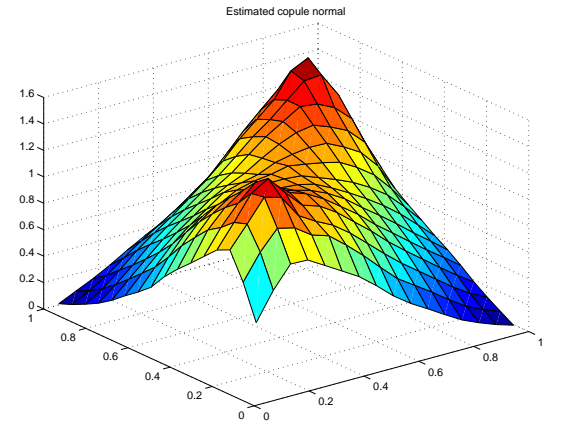
(a)



(b)



(c)



(d)

Figure 1. Estimation of the normal copula density of parameter 0.5 with $n = 2000$ (local thresholding): (a) true copula, (b) estimated copula with symmetrization, (c) estimated copula with periodization, (d) estimated copula with zero padding.

Copula		Method	Boundaries Handling		
$c(\cdot)$	par.		sym	per	ZeroPad
FGM	1.0	Local	0.007 (0.003)	0.079 (0.005)	0.129 (0.010)
		Block	0.006 (0.002)	0.077 (0.008)	0.141 (0.006)
normal	0.0	Local	0.002 (0.002)	0.0004 (0.0004)	0.122 (0.005)
		Block	0.002 (0.002)	0.0004 (0.0006)	0.105 (0.001)
normal	0.5	Local	0.031 (0.007)	0.161 (0.011)	0.179 (0.010)
		Block	0.032 (0.008)	0.154 (0.011)	0.202 (0.005)
normal	0.9	Local	0.156 (0.011)	0.391 (0.008)	0.418 (0.006)
		Block	0.140 (0.009)	0.381 (0.005)	0.491 (0.022)
Student	(0.5,1)	Local	0.326 (0.018)	0.460 (0.008)	0.544 (0.009)
		Block	0.324 (0.026)	0.458 (0.010)	0.585 (0.004)
Clayton	0.8	Local	0.075 (0.013)	0.225 (0.010)	0.252 (0.011)
		Block	0.095 (0.012)	0.216 (0.011)	0.279 (0.005)
Frank	4	Local	0.021 (0.006)	0.149 (0.015)	0.212 (0.015)
		Block	0.013 (0.006)	0.134 (0.009)	0.193 (0.006)
Gumbel	8.3	Local	0.701 (0.002)	0.849 (0.001)	0.866 (0.001)
		Block	0.698 (0.002)	0.852 (0.001)	0.878 (0.001)
Gumbel	1.25	Local	0.038 (0.010)	0.104 (0.005)	0.172 (0.009)
		Block	0.052 (0.007)	0.109 (0.004)	0.173 (0.004)

Table 1

Relative L_2 estimation error for $n = 500$

Copula		Method	Boundaries Handling		
$c(\cdot)$	par.		sym	per	ZeroPad
FGM	1.0	Local	0.0036 (0.0012)	0.0659 (0.0044)	0.0897 (0.0037)
		Block	0.0037 (0.0015)	0.0599 (0.0029)	0.1068 (0.0041)
normal	0.0	Local	0.0006 (0.0005)	0.0001 (0.0001)	0.0828 (0.0019)
		Block	0.0006 (0.0007)	0.0001 (0.0001)	0.0916 (0.0020)
normal	0.5	Local	0.0176 (0.0032)	0.1449 (0.0040)	0.1421 (0.0046)
		Block	0.0177 (0.0029)	0.1329 (0.0036)	0.1518 (0.0055)
normal	0.9	Local	0.1376 (0.0052)	0.3893 (0.0031)	0.4024 (0.0033)
		Block	0.1330 (0.0045)	0.3813 (0.0027)	0.4261 (0.0046)
Student	(0.5,1)	Local	0.2966 (0.0056)	0.4519 (0.0037)	0.5197 (0.0036)
		Block	0.2881 (0.0058)	0.4467 (0.0029)	0.5230 (0.0028)
Clayton	0.8	Local	0.0603 (0.0053)	0.2073 (0.0046)	0.2127 (0.0041)
		Block	0.0596 (0.0054)	0.1968 (0.0030)	0.2247 (0.0071)
Frank	4	Local	0.01208 (0.0032)	0.1244 (0.0047)	0.1186 (0.0043)
		Block	0.0075 (0.0017)	0.1137 (0.0035)	0.1218 (0.0048)
Gumbel	8.3	Local	0.6975 (0.0015)	0.8511 (0.0004)	0.8664 (0.0003)
		Block	0.6971 (0.0012)	0.8520 (0.0004)	0.8642 (0.0003)
Gumbel	1.25	Local	0.0240 (0.0041)	0.1022 (0.0030)	0.1392 (0.0029)
		Block	0.0336 (0.0042)	0.0988 (0.0026)	0.1503 (0.0038)

Table 2

Relative L_2 estimation error for $n = 2000$

Copula		Method	Empirical Loss Function		
$c(\cdot)$	par.		L_1	L_2	L_∞
FGM	1.0	Local	0.062 (0.014)	0.007 (0.003)	0.189 (0.051)
		Block	0.061 (0.011)	0.006 (0.002)	0.175 (0.047)
normal	0.0	Local	0.038 (0.017)	0.002 (0.002)	0.145 (0.062)
		Block	0.038 (0.018)	0.002 (0.002)	0.129 (0.058)
normal	0.5	Local	0.118 (0.012)	0.031 (0.007)	0.539 (0.066)
		Block	0.112 (0.016)	0.032 (0.008)	0.555 (0.051)
normal	0.9	Local	0.287 (0.026)	0.156 (0.011)	0.648 (0.020)
		Block	0.205 (0.021)	0.140 (0.009)	0.644 (0.018)
Student	(0.5,1)	Local	0.290 (0.022)	0.326 (0.018)	0.791 (0.026)
		Block	0.259 (0.018)	0.324 (0.026)	0.797 (0.035)
Clayton	0.8	Local	0.119 (0.014)	0.075 (0.013)	0.658 (0.051)
		Block	0.125 (0.018)	0.095 (0.012)	0.740 (0.040)
Frank	4	Local	0.129 (0.017)	0.021 (0.006)	0.329 (0.075)
		Block	0.092 (0.020)	0.013 (0.006)	0.321 (0.069)
Gumbel	8.3	Local	0.682 (0.015)	0.701 (0.002)	0.914 (0.001)
		Block	0.629 (0.012)	0.698 (0.002)	0.915 (0.001)
Gumbel	1.25	Local	0.099 (0.011)	0.038 (0.010)	0.625 (0.104)
		Block	0.105 (0.012)	0.052 (0.007)	0.749 (0.044)

Table 3

Relative L_1 , L_2 and L_∞ estimation errors for $n = 500$

Copula		Method	Empirical Loss Function		
$c(\cdot)$	par.		L_1	L_2	L_∞
FGM	1.0	Local	0.0448 (0.00821)	0.0036 (0.0012)	0.1414 (0.0382)
		Block	0.04887 (0.0096)	0.0037 (0.0015)	0.1463 (0.0527)
normal	0.0	Local	0.0181 (0.0087)	0.00063 (0.0005)	0.0673 (0.0332)
		Block	0.0190 (0.0092)	0.0006 (0.0007)	0.0669 (0.0284)
normal	0.5	Local	0.0830 (0.0078)	0.0176 (0.0032)	0.4374 (0.0465)
		Block	0.0923 (0.0104)	0.0177 (0.0029)	0.4089 (0.0673)
normal	0.9	Local	0.2048 (0.0160)	0.1376 (0.00522)	0.6400 (0.0114)
		Block	0.1622 (0.0113)	0.1330 (0.0045)	0.6389 (0.0106)
Student	(0.5,1)	Local	0.2159 (0.0107)	0.2966 (0.0056)	0.7712 (0.0110)
		Block	0.1955 (0.0095)	0.2881 (0.0058)	0.7669 (0.0133)
Clayton	0.8	Local	0.0862 (0.0068)	0.0603 (0.0053)	0.625 (0.0239)
		Block	0.1096 (0.0096)	0.0596 (0.0054)	0.6091 (0.0308)
Frank	4	Local	0.0983 (0.0131)	0.01208 (0.0032)	0.2635 (0.0569)
		Block	0.0702 (0.0096)	0.0075 (0.0017)	0.2508 (0.0608)
Gumbel	8.3	Local	0.6283 (0.0086)	0.6975 (0.0015)	0.9145 (0.0009)
		Block	0.6223 (0.0058)	0.6971 (0.0012)	0.9143 (0.0007)
Gumbel	1.25	Local	0.0720 (0.0075)	0.0240 (0.0041)	0.5377 (0.0568)
		Block	0.0721 (0.0085)	0.0336 (0.0042)	0.6688 (0.0421)

Table 4

Relative L_1 , L_2 and L_∞ estimation errors for $n = 2000$

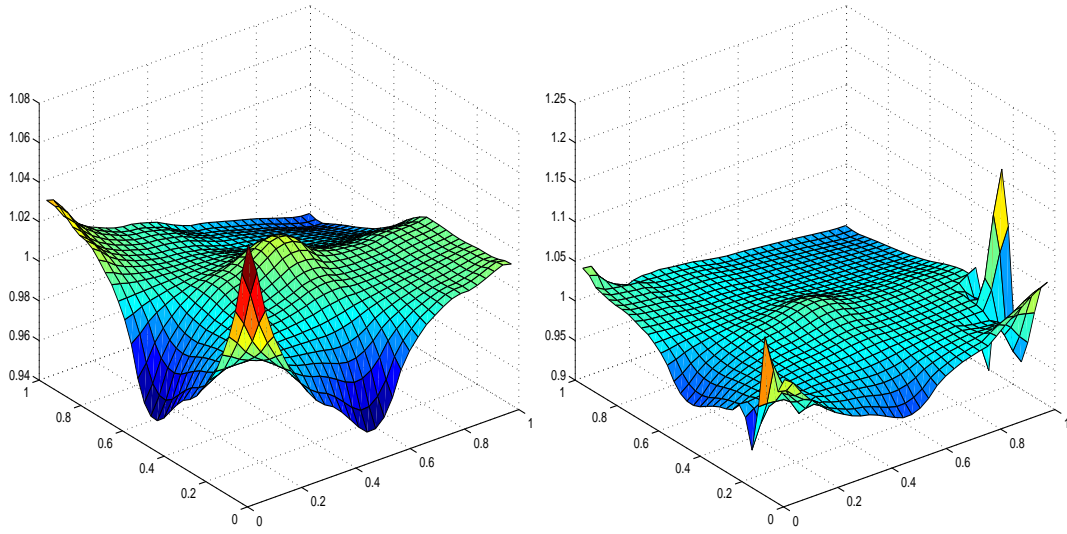


Figure 2. Brent/Cac: Block Thresh. Method (left) and Local Thresh. Method (right)

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	-0.01	0.0068	-0.01	0.0001	-0.01	0.0449
Gaussian	Local	-0.01	0.0080	-0.01	0.0002	0.01	0.0847
Student	Block	(-0.11,91)	0.0640	(-0.11,91)	0.0103	(-0.11,91)	0.6639
Student	Local	(0.07,40)	0.0226	(0.07,40)	0.0010	(0.02,100)	0.1279
Clayton	Block	0.01	0.0125	0.01	0.0002	0.01	0.0395
Clayton	Local	0.01	0.0135	0.01	0.0004	0.01	0.0942
Frank	Block	0.01	0.0103	0.01	0.0002	0.01	0.0467
Frank	Local	0.01	0.0115	0.01	0.0003	0.07	0.0825
Gumbel	Block	1.00	0.0093	1.00	0.0002	1.00	0.0462
Gumbel	Local	1.00	0.0106	1.00	0.0003	1.00	0.0963
All	Block	-0.01	Gaussian 0.68%	-0.01	Gaussian 0.01 %	0.01	Clayton 4.28 %
All	Local	-0.01	Gaussian 0.79 %	-0.01	Gaussian 0.02 %	0.07	Frank 7.98 %

Table 5

Brent/Cac: distances between the benchmarks and the parametrical families

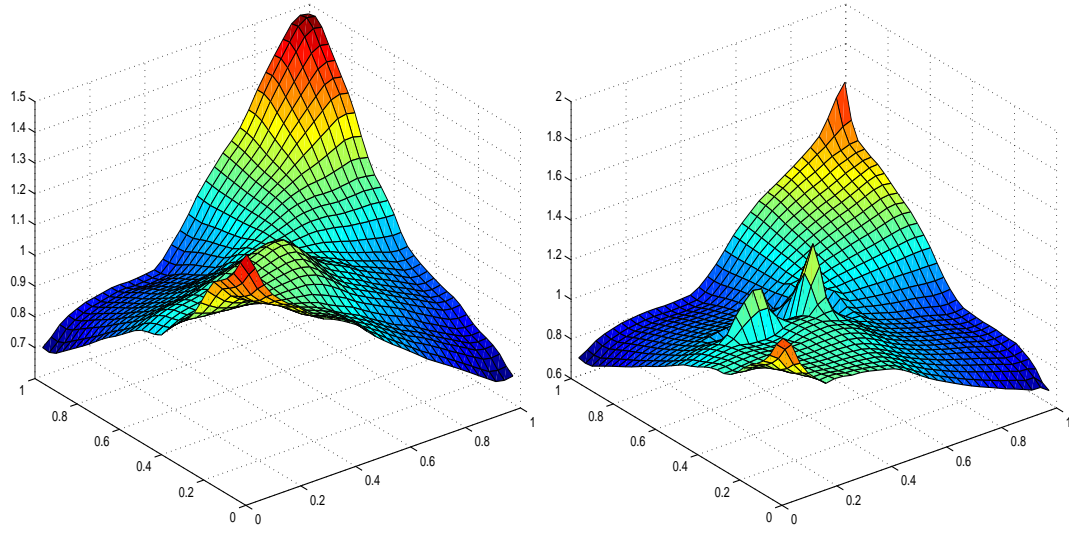


Figure 3. Brent/ExonMobil: Block Thresh. Method (left) and Local Thresh. Method (right)

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	0.15	0.0396	0.14	0.0030	0.10	0.1337
Gaussian	Local	0.14	0.0492	0.13	0.0041	0.10	0.1437
Student	Block	(0.14,37)	0.0376	(0.13,81)	0.0030	(0.08,61)	0.1329
Student	Local	(0.14,95)	0.0491	(0.13,95)	0.0041	(0.09, 80)	0.1411
Clayton	Block	0.15	0.0706	0.12	0.0099	0.05	0.1879
Clayton	Local	0.14	0.0799	0.11	0.0109	0.05	0.1967
Frank	Block	0.76	0.0301	0.83	0.0017	0.85	0.0957
Frank	Local	0.75	0.0393	0.80	0.0027	0.54	0.1355
Gumbel	Block	1.10	0.0436	1.07	0.0069	1.02	0.2309
Gumbel	Local	1.10	0.0529	1.06	0.0076	1.02	0.2298
All	Block	0.76	Frank 3.01 %	0.83	Frank 0.17 %	0.85	Frank 6.61%
All	Local	0.75	Frank 3.93 %	0.80	Frank 0.27 %	0.54	Frank 10.64%

Table 6

Brent/Exonmobil: distances between the benchmarks and the parametrical families

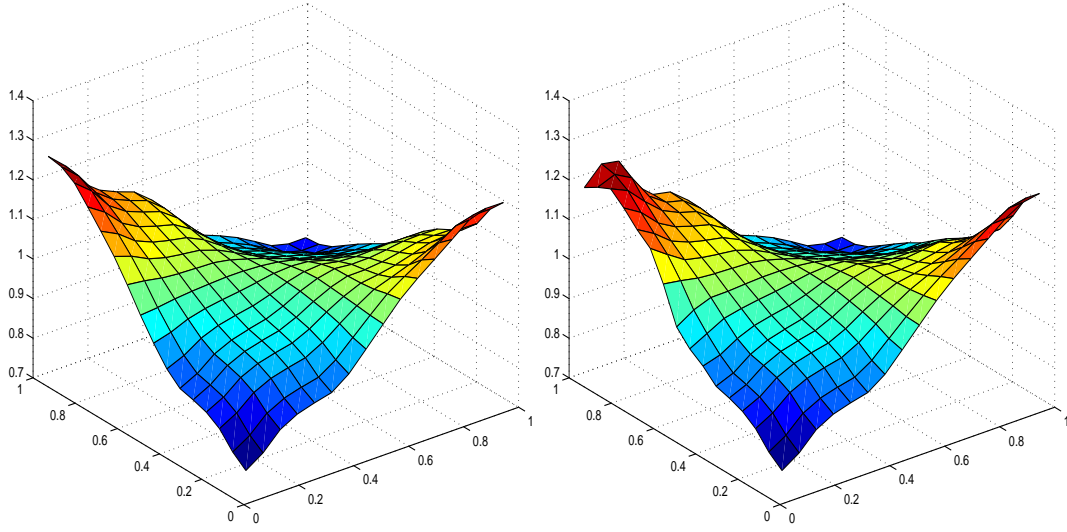


Figure 4. DowJones/Oncedcor: Block Thresh. Method (left) and Local Thresh. Method (right)

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	-0.11	0.0233	-0.10	0.0010	-0.07	0.0765
Gaussian	Local	-0.11	0.0243	-0.10	0.0011	-0.07	0.0765
Student	Block	(-0.11,61)	0.0233	(-0.10,61)	0.0011	(-0.06,61)	0.0859
Student	Local	(-0.11,80)	0.0239	(-0.10,80)	0.0011	(-0.06,63)	0.0859
Clayton	Block	0.01	0.0801	0.01	0.0104	0.01	0.2924
Clayton	Local	0.01	0.0805	0.01	0.0105	0.01	0.2924
Frank	Block	-0.57	0.0148	-0.56	0.0003	-0.50	0.0456
Frank	Local	-0.58	0.0155	-0.57	0.0004	-0.48	0.0433
Gumbel	Block	1.00	0.0755	1.00	0.0090	1.00	0.2316
Gumbel	Local	1.00	0.0760	1.00	0.0092	1.00	0.2316
All	Block	-0.57	Frank 1.48 %	-0.56	Frank 0.03 %	-0.50	Frank 3.69%
All	Local	-0.58	Frank 1.54 %	-0.57	Frank 0.03 %	-0.48	Frank 3.53%

Table 7

DowJones/Oncedcor: distances between the benchmarks and the parametrical families

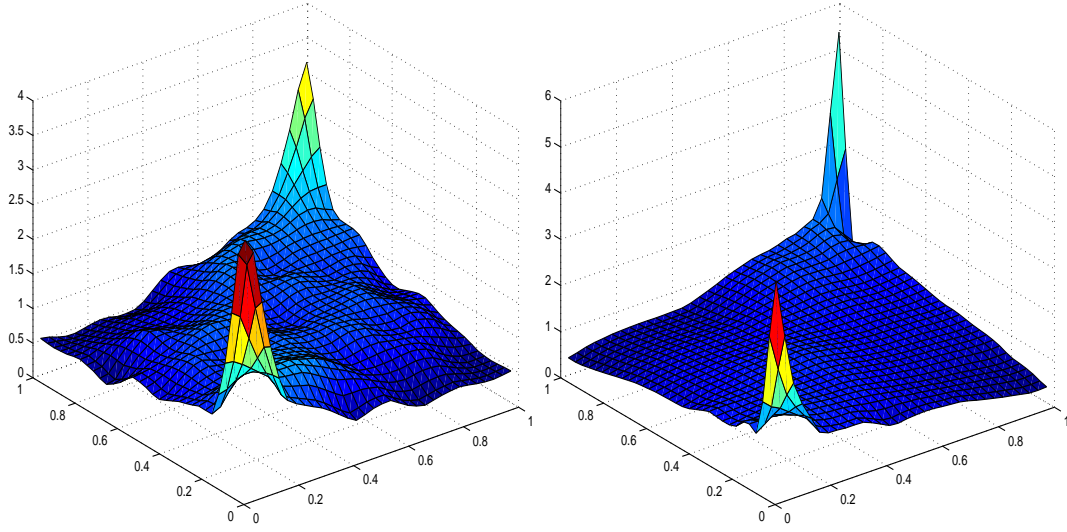


Figure 5. DowJones/Fste100uk: Block Thresh. Method (left) and Local Thresh. Method (right)

		$\hat{\theta}_1$	E_1	$\hat{\theta}_2$	E_2	$\hat{\theta}_\infty$	E_∞
Gaussian	Block	0.30	0.0976	0.33	0.0202	0.20	0.4191
Gaussian	Local	0.26	0.0699	0.32	0.0234	0.11	0.2785
Student	Block	(0.28,8)	0.0755	(0.29,8)	0.0127	(0.18,11)	0.3027
Student	Local	(0.17,12)	0.0846	(0.17,6)	0.0265	(0.12,20)	0.3748
Clayton	Block	0.40	0.1064	0.36	0.0318	0.26	0.4565
Clayton	Local	0.31	0.0978	0.33	0.0401	0.11	0.3465
Frank	Block	1.58	0.1094	1.88	0.0333	0.57	0.4366
Frank	Local	1.38	0.0687	1.73	0.0401	0.79	0.2762
Gumbel	Block	1.19	0.1081	1.17	0.0414	1.09	0.4427
Gumbel	Local	1.18	0.0782	1.18	0.0282	1.06	0.3866
All	Block	(0.28,8)	Student 7.55 %	(0.29,8)	Student 1.15 %	(0.18,11)	Student 10.62%
All	Local	1.38	Frank 6.86 %	0.32	Gaussian 2.12 %	0.79	Frank 19.56 %

Table 8

DowJones/Ftse100uk: distances between the benchmarks and the parametrical families